**The Thesis Committee for Weitong Sun**
**Certifies that this is the approved version of the following Thesis:**

# A Scenario Management Platform that Incorporates Statistic and Simulation for Unconventional Field Development

**APPROVED BY**
**SUPERVISING COMMITTEE:**

Kamy Sepehrnoori, Supervisor

Wei Yu

# A Scenario Management Platform that Incorporates Statistic and Simulation for Unconventional Field Development

**by**

**Weitong Sun**

**Thesis**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Master of Science in Engineering**

**The University of Texas at Austin**

**May 2019**

# Dedication

To my parents, **Zhinan** and **Jun**

## Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Kamy Sepehrnoori, for his insightful ideas, continuous guidance, and encouragement throughout the past three years. His immense knowledge in reservoir simulation has always guided me to explore this area.

I am grateful to Dr. Wei Yu, for his time and effort to review my thesis. I benefit so much from the invaluable comments and feedback he provided. The technical discussion with him deepened my understanding of reservoir simulation. I also learned software skills and writing skill from him.

I would like to thank Silpakorn Dachanuwattana, for his time to discuss history matching with me. His excellent work is the basis of my research.

I will also extend my thanks to all my friends and officemates, for the great time I spent with them at UT.

I appreciate the financial support from Statoil and the members of Reservoir Simulation Joint Industry Project (RSJIP) at the Center for Petroleum and Geosystems Engineering at The University of Texas at Austin.

Finally, I would like to express my deepest gratitude to my parents, for their continuous love and endless support. Their encouragement constantly motivated me in my work and life.

My life is truly blessed to have you all.

# Abstract

## A Scenario Management Platform that Incorporates Statistic and Simulation for Unconventional Field Development

Weitong Sun, M.S.E.

The University of Texas at Austin, 2019

Supervisor: Kamy Sepehrnoori

Producing from shale formations has been made profitable because of technological advancements. However, the complexity and uncertainties of the unconventional reservoir make it hard to estimate the assets and maximize the value. Reservoir simulation is a powerful tool to estimate the performance of reservoir but calibrating models and optimizing the development plan can take lots of human efforts and computation time, especially when we need multiple models to stress the uncertainties. Many of methods have been developed to improve the efficiency of simulations and reduce the number of simulations needed. There are also analytical packages to help understand the results of simulations from the statistical point of view and build economic models. Thus, an efficient way to incorporate the necessary tools and methods from different sources can be helpful for the decision-making process.

The designed scenario management platform can help to understand the uncertainties and to make decisions by analyzing the possible scenarios and correlated data. Connected by the data structure management system, the system is equipped with four

primary modules, sampling, modeling, calculation interfaces, and visualization tools. The modules can work separately to carry out works like a predictive statistical model, lunch a batch of simulation according to the template and uncertainties, sampling improve the model or according to a distribution, access the model and presenting results. They can also be used together to do more comprehensive work like history matching and well spacing.

This thesis presents a few of the technics that are implemented in this platform that can be helpful to understand the uncertainties. We also show some of the applications enabled by the modules of this system and some of the visualization ideas to diagnose the models.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1: Introduction

## 1.1 Digging Values from Fractured Reservoirs

Production from shale formations has been made profitable because of technological advancements of horizontal drilling and multi-stage hydraulic fracturing. However, the profitability of unconventional development depends on the quality of the knowledge we obtained in a short period and the decisions we made upon that knowledge.

Reservoir simulation is widely used in the oil and gas industry to evaluate the assets and make development plans. However, the gathered static data such as geological information, seismic interpretation, and petrophysical data have inherent uncertainties that can compromise the credibility of simulation results; the simulator calculated production profile might be different from that of actual production history.

History matching (HM) is a process that can calibrate the reservoir model with production history. Besides increasing the ability to have a reliable production forecast, we can characterize the reservoir better, which is especially useful for the further development of unconventional reservoirs. However, calibrating models for unconventional reservoirs can take huge human efforts and expensive computation time.

The fractures in unconventional reservoirs are the main reason we need to spend more time to do a single simulation and run more simulations for history matching. Fracture geometries have a significant influence on production performance but are not easy to capture in detail. Since the property of the fractures can be distinctively different from that of the matrix, avoiding the numerical issues for simulations can make the computation expensive. At the same time, extra uncertain properties need to be added to describe the distribution of the fractures, which will increase the dimension of uncertain space and the efforts necessary to explore it.

Optimizing the development plan for the unconventional reservoirs is also a demanding job when the uncertainties we quantified are taken into account. Given such uncertainties underground, the production estimation shall be in the form of distribution.

The combination of potential reservoir realizations and the development plan may lead to a large number of simulation needed.

The motivation to estimate the assets and maximize the value of fractured reservoir is enormous despite the challenges above. Many tools from different sources are developed to improve the efficiency of simulations and reduce the number of simulations needed. Fracture models like embedded discrete fracture model (EDFM) can reduce the time needed for simulation. History matching workflows designed for unconventional reservoirs can reduce the number of simulations needed and quantify uncertainties at the same time. There are also geomechanical models developed to understand the properties of fractures.

Utilizing newly developed tools and emerging methods for different problem-solving workflow is a tedious job. As we may need a large number of simulations to promote our decision, automating the process and analyzing results can be a large workload. Thus, a fast and efficient system to quantify uncertainties and test development plan can be a helpful tool for the decision-making process.

### 1.2 OBJECTS OF THIS RESEARCH

Based on the problems described above, the objectives of this thesis are:

1. Design typical workflows that can help understand and manage the uncertainties for the development of unconventional reservoirs.

2. Use different methods to reduce the computation time needed for simulations and the number of simulations needed.

3. Find the standard components in the workflows and modularize those functions for universal usage. Implement each module with the designed interface that connects them.

4. Extract the critical concept in each module and do the modification to make it possible to incorporate new tools into the workflows easily.

5. Provide examples that manage the unconventional reservoirs using our platform.

## 1.3 Brief Description of Chapters

In Chapter 2, a literature review of some technics used to manage the unconventional reservoirs is presented.

In Chapter 3, first we go through the structure of our scenario management platform designed to help manage the unconventional field, then the methodology and formulations that empowered each module of our platform, and finally the workflows that solve typical problems during the management of unconventional reservoir using our platform.

Chapter 4 introduces our practice to utilize artificial neural networks as a predictive model, followed by several tools and strategy that helps to improve the models for specific purposes.

In Chapter 5, a shale gas example is used to demonstrate the history matching and well-spacing management workflow.

In Chapter 6, a gas condensate example is presented to demonstrate the workflow to handle history matching problem with more uncertainties and complex fluid properties.

In Chapter 7, the history matching results of Chapter 6 are analyzed with interactive parallel coordinates plots which are designed for the visualization of higher dimensional data and models.

Finally, in Chapter 8, all the studies are summarized, and recommendations for future studies are given.

# Chapter 2: Literature Review

Uncertainties behind the unconventional reservoirs make the field management resources-demanding, especially when the simulation is the primary tool to estimate the production. Assisted history matching is widely used to calibrate the reservoir model and quantify the uncertainties underground. To reduce the resources needed, fracture models are developed to accelerate the simulation for fractured reservoirs; proxy models are used to generalize the knowledge from the simulation. After history matching, the uncertainties captured can be used to adjust the development plan to maximize the expected value. In this chapter, we present some of the progress has been made improving the effectiveness and efficiency of those process.

## 2.1 Uncertainties of Fractured Reservoir

Studies on fractured reservoirs were triggered by its economic potential. Many hydrocarbon resources in tight reservoirs are regarded commercially recoverable, and hydraulic fractured horizontal wells are regarded a good way to promote the production (Jennings et al., 2006; Shaoul et al., 2007; Medavarapu et al., 2012; Murtaza et al., 2013). However, modeling the fractured reservoir is a challenging job considering the information we can collect.

The uncertain nature of the fractured reservoir makes it hard to assign properties to the model. For unconventional reservoirs, the challenges of property characterization exist in measurements (Forsyth et al., 2011; Sakhaee-Pour, 2012; Frash et al., 2014) and analysis (Schuetter et al., 2015; Steiner et al., 2015). The permeability of shale gas reservoirs could range from 100 to 10000 nd, and the measurement is hard for nanoscale pore throats (Sakhaee-Pour, 2012). For this kind of reservoir, the shape and conductivity may be the dominant factor for fluid flow. However, the fracture networks underground are hard to measure and even harder to model. A typical way to model hydraulic fractures is through

the bi-wing fracture where the planner fracture propagates in the direction that is perpendicular to the least principal stress (Perkins and Kern, 1961; Nordgren, 1972). The properties for this simplified model like fracture half-length, fracture height and fracture conductivity (Yu and Sepehrnoori, 2014, 2018) are not natural properties that can be measured and thus are supposed to be regarded uncertain too.

The uncertainties of the reservoir properties have an unavoidable effect on production prediction and economic analysis. Nonetheless, they can be alleviated or even quantified by assisted history matching.

## 2.2 ASSISTED HISTORY MATCHING

History matching calibrates the reservoir flow model by minimizing the discrepancies between the observed and simulated production data (Ertekin, 2011). The history matched reservoir models are usually believed to represent the actual reservoir, and therefore are used for predictions and management. Since the history matching is generally considered as an ill-posed inverse problem, we want not only multiple solutions, but also the probability of the solutions. This makes manual history matching practically impossible considering the multi-dimensional uncertainties we are facing.

Various assisted history matching (AHM) workflows have been promoted to find multiple history matched realizations and quantify the uncertainties of the reservoir. The main branches for AHM include optimization-based methods, Ensemble Kalman Filter (EnKF) and Bayesian methods (Oliver and Chen, 2010).

The optimization-based methods use optimization algorithms to minimize the matching error gradually. Some notable practice includes adjoint and gradient method (Gang and Kelkar, 2008), genetic or evolutionary algorithm (Yin et al., 2011; Zhang et al., 2013), particle swarm (Vazquez et al., 2015) and gauss-Newton (Gao et al., 2016; Chen et al., 2017). This kind of method usually cares more about the quality of a small batch of reasonable solutions and requires low computation cost. The tradeoff is that the solutions ensemble we get is primarily affected by the optimization algorithm (Erbas and Christie,

2007) and the actual probabilistic distribution can not be gathered directly (Goodwin, 2015).

EnKF is a proven method for the conventional reservoir where heterogeneities introduced a large number of not-so-influential uncertainties. It can also provide a posterior distribution, but there is still a question whether they can model the non-linear relationship between input and response parameters and reproduce the true distribution (Emerick and Reynolds, 2012, Goodwin, 2015).

The Bayesian method can be applied on HM for the solution ensembles that represent the posterior distribution of uncertainties in the reservoir. The Markov Chain Monte Carlo (MCMC) is one popular Bayesian method to sample a posterior distribution, yet it usually takes $10^5$-$10^6$ transitions to obtain a stable chain that can represent the posterior distribution. It is not practical to do so many simulations despite its robustness. Thus, rather than directly running so many simulations, proxy models are usually utilized to take the place of actual simulations. The proxy-based MCMC workflows for AHM are used in the conventional reservoirs (Slotte et al., 2008; Goodwin, 2015) and unconventional reservoirs (Wantawin et al., 2017a; 2017b).

### 2.3 EDFM

Fluid flow in an unconventional reservoir is usually dominated by the properties of fractures and its interaction with matrix rather than small scale heterogeneity of the matrix. Thus, it is important to have a fracture model that can capture the behaviors around the fractures (Lemonnier and Bourbiaux, 2010).

Methods like dual porosity and dual permeability (DPDP) approach (Warren and Root, 1963; Blaskovich et al., 1983), and discrete fracture models (DFM) were proposed (Noorishad and Mehran, 1982; Baca et al., 1984; Kim and Deo, 2000; Karimi-Fard and Firoozabadi, 2003). DPDP is not suitable for large scale fractures (Karimi-Fard et al., 2004) or accurately evaluating the fluid transfer between matrix and fractures (Karimi-Fard et al., 2004; Monteagudo and Firoozabadi, 2004) which makes it not suitable for the hydraulic fractured reservoirs. DFM is developed to solve this problem, but it also introduces a large

number of small grids around the fracture, which increases the complicity of gridding and cost for computation.

The Embedded Discrete Fracture Model (EDFM) was proposed to overcome this issue by introducing additional fracture grids aside the original model and modify corresponding non-neighboring connections and properties to play the role as fractures. It can be implemented for the in-house reservoir simulators (Moinfar et al., 2014; Shakiba et al., 2015; Yu et al., 2018a, 2018b) and as a preprocessor for commercial reservoir simulators (Xu et al., 2017a, 2017b, 2018; Dachanuwattana et al., 2018; Tavassoli et al., 2018; Fiallos Torres et al., 2019a, 2019b; Yu et al., 2019; Eltahan et al., 2019). The EDFM preprocessor was used to model any complex natural and hydraulic fractures for all the simulations in this thesis.

## 2.4 PROXY MODEL

A proxy model is a mathematical model used to approximate the response of an unmeasured point cheaply based on measured points. It is also called a surrogate model or a response surface model in some literature. Different mathematical models have been applied as proxy models on different problems in the petroleum industry. Popular models used in AHM include but not limited to polynomial regression (Zhang et al., 2013; Wantawin et al., 2017a, 2017b), K-nearest neighbor (Yu et al., 2018c), kriging or Gaussian processes model (Landa et al., 2003; Hamdi et al., 2017), support vector regression (Guo et al., 2017; Dos and Reynolds, 2019), and artificial neural network (Ramgulam et al., 2006; Shahkarami et al., 2014; Siripatrachai et al., 2014). Despite its popularity, the proxy models are suffering from problems like accuracy, overfitting, and computation time. The implementation of the proxy model may affect the result.

The booming developments in machine learning society provide easy access to the most advanced tools. Open sources tools for the artificial neural network like TensorFlow and Keras have been used in data interpretation (Dramsch and Luthje, 2018; Wu et al., 2018) and value prediction (Luo et al., 2019; Temizel, 2019). With the tools ready-to-go,

we can easily iterate through different methods for our problem without diving into the details of the implementation details.

### 2.5 UTILIZING THE RESULTS OF HISTORY MATCHING

The knowledge can be gained depends on the history matching method used and the steps after that. The primary target is usually the calibrated model for the further development plan. Modern AHM technics can provide multiple history matched solutions and are widely used to calibrate the reservoir model and quantify the uncertainties of the reservoir (Goodwin, 2015). But most publications only provide limited studies beyond the history matching.

The history matched models are usually used directly when the further simulation is needed. 30-year production for all the history matched scenarios can be simulated to estimate ultimate hydrocarbon production (Wantawin et al., 2017a, 2017b; Yu et al., 2018), and to calculate net present value (NPV) when economic environments are assumed. When more complex simulations are needed, one or several history matched models can be selected and used to optimize the well placing plan (Özdogan and Horne, 2004) or enhanced oil recovery (EOR) plan (Alfarge et al., 2018).

The quantified uncertainties are usually presented as a probability distribution function (PDF) to increase the knowledge about the assets. For unconventional reservoirs, the PDF for the shape and distribution of hydraulic fractures can help to understand the quality of completion (Wantawin et al., 2017; Yu et al., 2018c). Since tools like MCMC can sample the posterior distribution of uncertainty, with the help of the proxy model, P10-P50-P90 of ultimate production can be provided in the fashion of posterior distribution.

Optimization under uncertainty requires large amount of computation efforts since field development options need to be benchmarked against a reasonable number of subsurface scenarios (Schulze-Riegert et al., 2017). Different workflows (Wang et al., 2012; Hanea et al., 2015) have been suggested to optimize the development plan when the uncertainties are provided. The workflow to perform optimization after history matching is also promoted (Schulze-Riegert et al., 2017). In that workflow, the uncertainties are

evaluated on the basis of a representative subset of history matched scenarios, but the scenarios are chosen base on probability distribution of original gas in place. A better way to honor the multi-solution feature of history matching may be more feasible.

# Chapter 3: Structure of the Platform and the Methodologies

This chapter starts with a brief introduction to the main structure of the platform. Subsequently, the methodology and ideas behind the primary functions of each module within our platform are presented. After that, we walk through a few example workflows enabled by the platform that can help us with the management of the unconventional reservoirs.

## 3.1 Structure of The Platform

The platform is designed base on the experience gained from history matching and well spacing projects of the unconventional reservoir. Each module of the platform is one of the vital functions that is frequently used in typical workflows.

### 3.1.1 Decomposing the Problem

A typical practice for the unconventional reservoir to evaluate the stimulation effectiveness is the history matching and estimated ultimate recovery (EUR) prediction. In common practice, the two processes are usually done simultaneously. An example history matching workflow is shown in Figure 3.1. In this workflow, the history matching is done with iterative response surface method, and the probability forecasting is done with the proxy based MCMC method. The two processes are independent of each other, except that the probability forecasting needs the uncertainties quantified by the history matching process as the input. The uncertainty is the critical information we care about as we want to quantify it and make decision utilizing it. On the one hand, there are various methods to quantify the uncertainties with history matching or get it from other sources. On the other hand, we can use the uncertainties quantified in other works like well spacing and enhanced oil recovery (EOR) planning. Thus, uncertainty quantification and utilization will be two independent problems we care about. Lucky for us, the key components to solving those two different problems can be very similar when using some of the methods. In this study, we merely explore the methods that utilize the results of ensemble simulations.

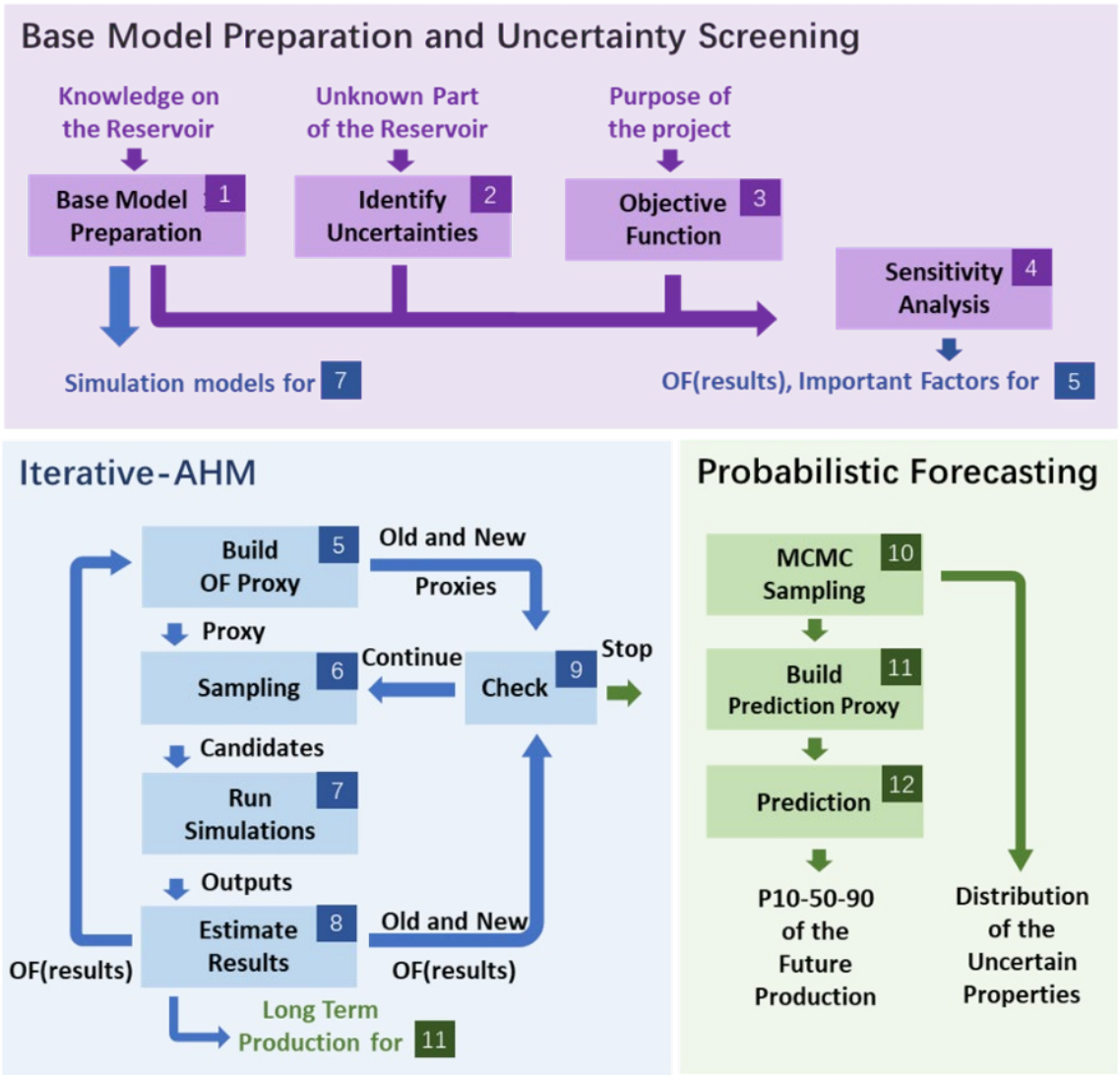



Figure 3.1: An interactive assisted history matching workflow to study the uncertainties.

### 3.1.2 Modules of the Platform

For the platform, each module represents a necessary function for the workflow to study the uncertainties of the unconventional reservoir. In this study, the primary method to deal with the uncertainties is by evaluating possible scenarios of the uncertainty and provide the understanding of the problem with the obtained results. A typical workflow

usually includes designing the scenarios according to the need, evaluating the scenarios at a reasonable cost, and evaluating the effect of the uncertainties on a target problem.

In this study, the uncertainties are dealt like probability distribution function (PDF) in high dimensional spaces. Each uncertain property is viewed as a variable, and all the combined uncertain properties form a high dimensional space. A possible realization of the reservoir under uncertainty is equivalent to a point in this high dimensional space. The probability of each point in this uncertain space to be the real reservoir forms the PDF of this uncertain space. The objective is to get this PDF and understand its influence on the possible development plan.

Designing the scenarios under uncertainties is viewed as sampling from the uncertain space. Thus, the sampling module serves the purpose to design scenarios for specific goals like building a better surrogate model or getting a wanted distribution. A common approach to evaluate the scenario is through the numerical simulation or some equations. The calculation interface module is designed to incorporate those into the workflow easily. The module of statistic models is used to help evaluate the scenarios at a considerably lower cost. The platform also provides some visualization design that helps us monitor the process and present the results. The main structure of the platform and the primary functions of each module are shown in Figure 3.2.

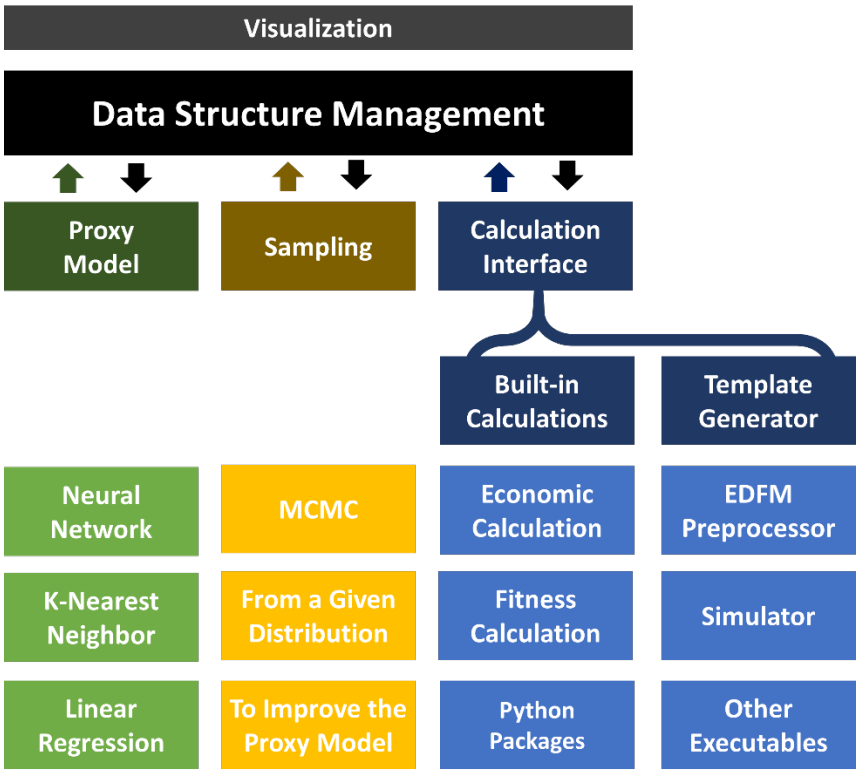



Figure 3.2: The structure of the platform and typical functions of each module.

## 3.2 Sampling Module

The sampling module is used to design the scenarios for different purposes. In this section, we present four primary goals and introduce the strategy we use when we have multiple goals during the sampling process.

### 3.2.1 Objectives of Sampling

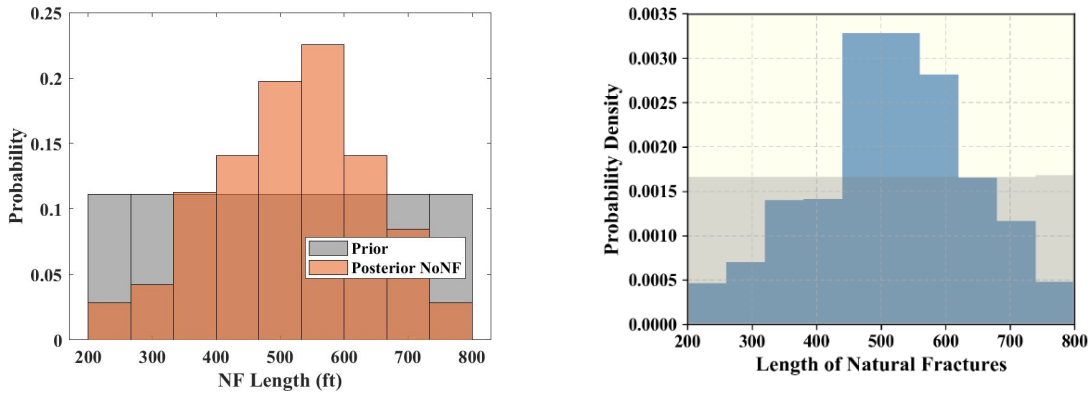
The objective of sampling decides how we perform the implementation. Moreover, things become complicated when we need to achieve multiple goals at the same time. Each chosen experiment ran by the simulator can be viewed as a point in the high dimensional space formed by different uncertain variables. Each point in this space has a corresponding tensor representing the objective function.

#### *3.2.1.1 Base on a Sample*

The most common reason we need to create the scenarios is that we need the realizations of the uncertainties to run the simulation and get reliable estimations of the assets. The uncertainties are usually presented as a distribution.

When we have a probability distribution function that can describe the distribution, it is easy to generate a wanted number of the sample from this distribution. However, sometimes what we can work on is a bunch of samples that cannot be described by simple distribution functions.

The easiest method to do this is to divide the range into small bins and count the cumulative number of the data in each bin and generate a corresponding number of the data by portion. We notice that the size of bins will affect the shape of PDF a little and thus need special care when the sample size is small. Figure 3.3 shows one of the examples our platform generates the samples from another sample.




(a) Old sample (b) New sample

Figure 3.3: Generate (b) new sample base on an (a) old sample from another source.

#### *3.2.1.2 Quantify Uncertainties with MCMC*

Another popular objective is that we want to get the posterior distribution of the uncertainty. One of the tools we can utilize for that is Markov Chain Monte Carlo

(MCMC). A Markov Chain that reaches equilibrium can be constructed to obtain a sample of desired distribution.

In our platform, we use the Metropolis-Hasting method with an evaluation function that is built upon the evaluated samples. The chain element generated in step i is marked as $\theta^i$, and is a multi-dimensional point in the uncertain space that represents a realization of the reservoir model. Since the random initial point $\theta^0$, each generated point $\theta^i$ is followed by a random proposal point $\theta^{i*}$ within a distance to $\theta^i$. We evaluate the proposal point using Eq. 3.1. Where E in this function is the objective function of interest and $\sigma^2$ is the variance in Gaussian distribution. If $P(\theta^i \to \theta^{i*})$ is compared to a random number between 0 and 1, and the proposal $\theta^{i*}$ is accepted when it is larger than the random number. The $\theta^{i+1}$ is set to be the proposal point if the acceptation rule is satisfied. Otherwise, the proposal $\theta^{i*}$ is rejected and $\theta^{i+1}$ is set to be $\theta^i$. The chain we generated is regarded as a sample with the posterior probability distribution after the chain is long enough to converge.

$$P(\theta^i \to \theta^{i*}) = e^{\frac{E^2(\theta^i) - E^2(\theta^{i*})}{2\sigma^2}}, \tag{3.1}$$

#### *3.2.1.3 Improve the Proxy Model*

Since simulation is computationally expensive and takes considerable time, we sometimes use a proxy model to help evaluate the scenarios. Thus we need to make the proxy as accurate as possible. Besides choosing proper models, it is crucial to choose the scenarios to be used to build the proxy model.

The idea behind this part is more intuitional than mathematical. Since the problems are unknown and not unique, no plausible universal theory is available to tell us where we shall sample to improve the proxy model. The intuitive idea is that the prediction is more accurate around the place with more samples that have been evaluated by the simulator. At the same time, the forecast will be less reliable when we have less evaluated scenarios around. If the accuracy of the overall performance of the proxy model is our priority, then it will be better if we can have points remotely distributed in the uncertain space.

#### *3.2.1.4 Optimize the Scenarios*

Another typical problem we are facing is that we want to get scenarios that minimize an objective function. This can be a complicated process and has many approaches, yet for the platform, the iterative response surface method is chosen to make the process non-intrusive method. For this method, a proxy model is built to evaluate the scenarios at the begging of each iteration. The proxy-evaluated results are then used to sample the scenarios to be evaluated by the simulator. Then, the simulator-evaluated scenarios will be used to update the proxy model at the end of this iteration. A simple way to sample for this purpose would be directly rank the scenarios and select them to the rank.

#### *3.2.1.5 Joint Objectives*

In the real world, one objective at a time may not be good enough, and sometimes we want to achieve multiple goals at the same time. Take history matching problem as an example; the traditional goal of history matching is to find the region of the uncertain space with the lowest matching errors.

An efficient and effective sampling method for our workflow will be capable of optimizing the objective function fast while providing enough scenarios to build a good proxy model for forecasting step, which brings in the three goals.

(1) From the optimization point of view, we need a sampling method to choose the scenarios that have less error but cover more area at the same time. For the proxy-based workflow, we can feed the simulator scenarios that minimize the objective function according to the prediction of the proxy model. For most proxy models, those predict-to-be less-error points clustered around the old points usually leads to less simulation error.

(2) On the other hand, to avoid trapping in a local minimum when the preferable goal is the global minimum, we need to explore the unknown area at the same time, which means we cannot only focus on a small region. Also, we want to know about the region that has less error even if they are not the global minimum as they are also a possible solution with fewer possibilities.

(3) Besides helping to search for the solutions in the optimization step, we will also use our proxy models built to quantify the uncertainty. To have an accurate overall prediction in the uncertain space, we need to have points distributed everywhere. To have a better prediction around the solutions, we need to have more experiment points near the solution regions.

Those goals are contradicting each other given that we are not going to test too many scenarios due to the computational cost of the simulation. Thus the idea of sampling unit and sampling strategy are introduced.

### 3.2.2 Sampling Unit

The idea behind the sampling unit is to divide a large problem into small pieces. A practical method to divide the problem is to sample by iteration and take advantage of the updated knowledge at the end of each iteration. It is easy to apply different sampling algorithms to focus on one goal at a time for each iteration, yet it may add more cost. The recommended approach applied in the later field case is by combing the sampling algorithms and use a governing equation to decide the samples to be chosen from.

We introduce a concept of the sampling unit which uses a specific algorithm to generate samples or choose samples from a population. The sampling units are object-oriented, and the way to implement may not be identical as long as it serves the object. Thus, the sampling process with a single objective, like the ones we mentioned in the previous section, is by nature a sampling unit. Here we introduce another few sampling units that can be used to assist the process.

#### *3.2.2.1 Diverging Sampling Unit*

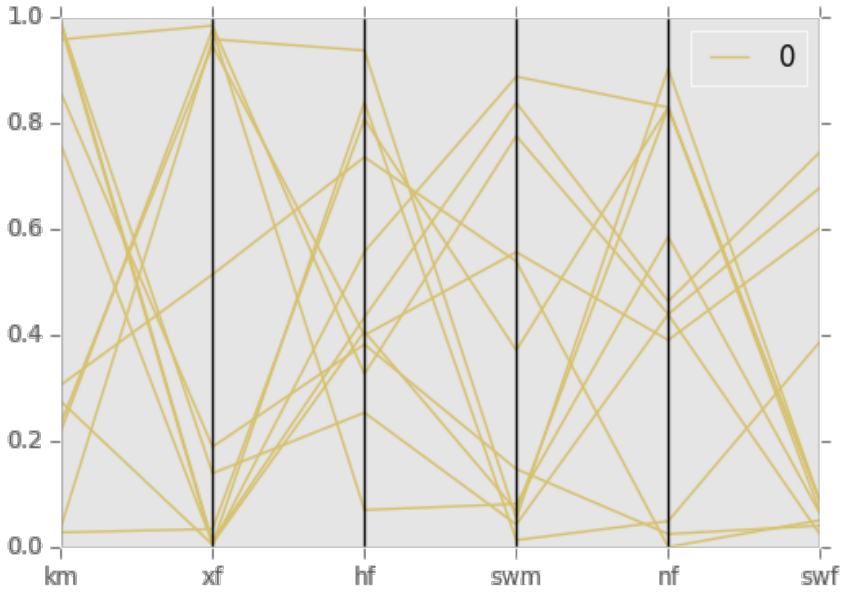
A diverging sampling unit pulls a set of samples that are far away from each other in the high dimensional space and thus controls the diversity.

In this study, a Dijkstra-like algorithm is promoted to help find a set of points scattered in the space from a large number of given points. The algorithm is designed to provide considerable good filtering rather mathematically scattered set, balancing the

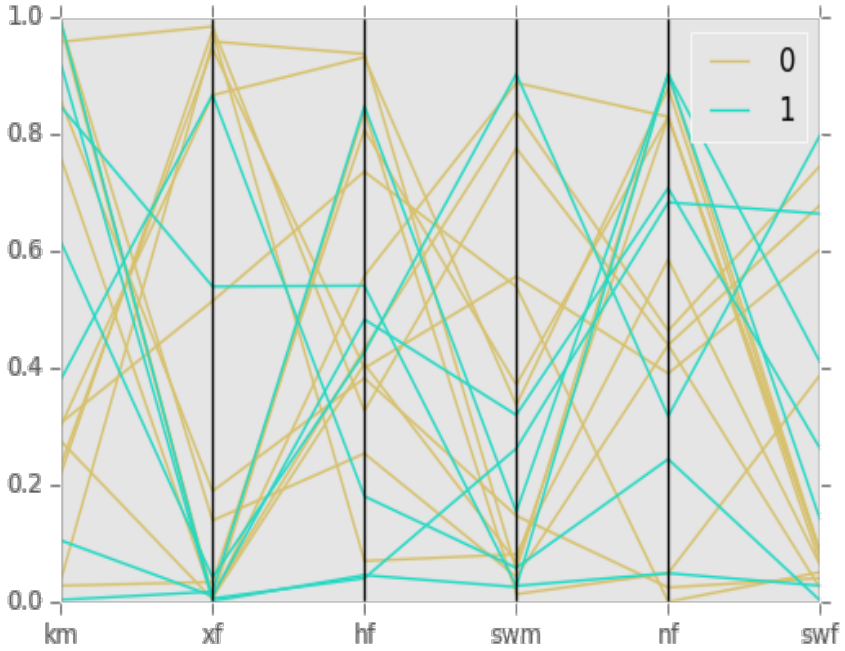
efficiency and performance. The steps of the Dijkstra-like Remote algorithm [RM(Si,Sb,Nr)]are listed below:

1. Input the sample list (Si) with a population of Ni, base sample (Sb) with a population of Nb and the objective count of result samples (Nr). Create a list (Ld) with the length Si and an empty list Sr.
2. Calculate the distance of the first point to every point in Sb and record the minimum distance in Ld.
3. Do the calculation for the rest of the points in the Sb and record the results in Ld. Ld is the list to record the smallest distance.
4. Add the point (Pt) with the lowest distance to Sr according to Ld. Delete this point for both Si and Ld.
5. Calculate the distance of the points in Si to Pt and update Ld.
6. Repeat step 4 and 5 until we have Nr points in the Sr.
7. Sr is to be output as the sample.

For this method, the computational cost can be high if (Nr+Nb)×(Nr+Ni) is large and we would like to limit the Nr to a considerably small number. Also, Sb is not a necessary input if diverging is not the only purpose. We may only focus on this iteration rather than consider the effect of previous ones. An example to add points by batch with diverting sample unit is demonstrated in Figure 3.4. Lines labeled 0 are the old samples compared to the lines labeled 1. As can be seen, new samples are covering the space that is less covered by the old samples.

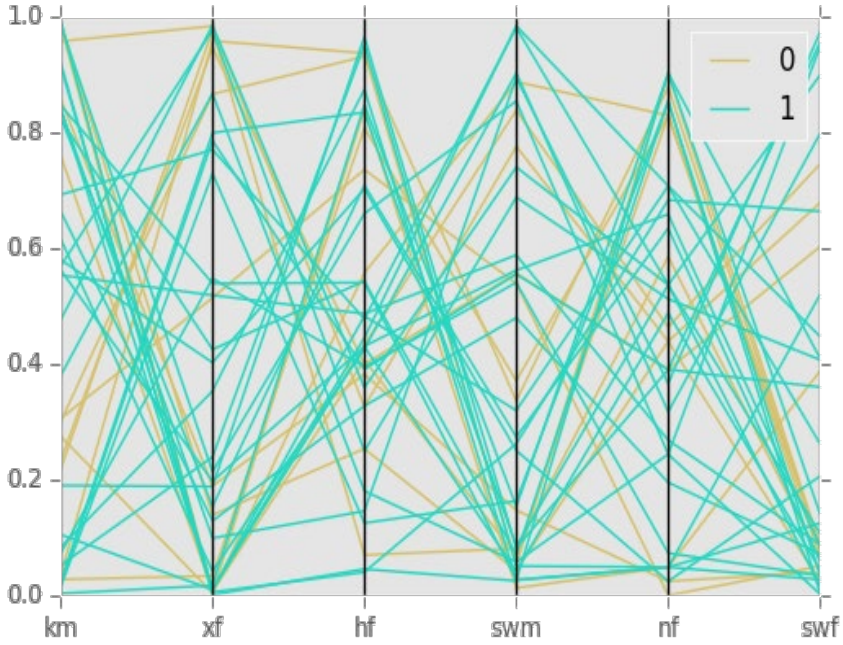



(a) The first batch of Sample




(b) The second batch of sample

Figure 3.4: Adding (a) the first, (b) the second, and (c) the third batch of samples with the diverging unit.




(c) The third batch of sample

Figure 3.4 continued

#### *3.2.2.2 Optimum Sampling Unit*

An optimum sampling unit selects a set of samples that are optimal according to an objective function. For our platform, the objective function can be either directly calculated or evaluated by the proxy model.

#### *3.2.2.3 Initial Sampling Unit*

Since the previous two sampling units are all sampling from an existing sample, we need a method to provide the initial sample. A typical example of the initial sampling unit is random sampling which randomly generates a wanted number sample. Other sampling generated by the methods in section 3.2.1 can also be used as the initial sampling unit in the process.

### 3.2.3 Sampling Strategies

The sampling strategy is the plan that decides which sampling unit to use and how to arrange the sequence. The strategy will be different according to the goals, and an example is shown in section 6.3.

Stage 1 (more global):

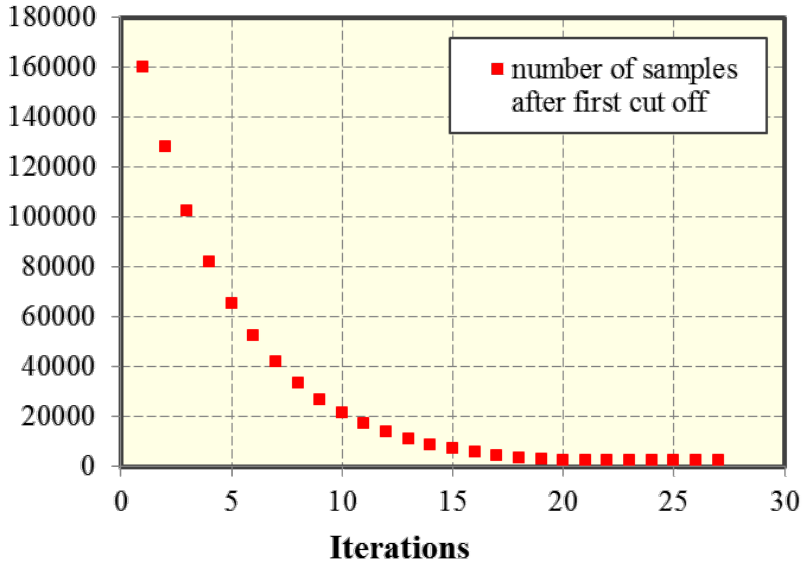
1. More exploration of the unknown space of uncertain variables
2. More points on the space closer to good estimates

Stage2 (more local):

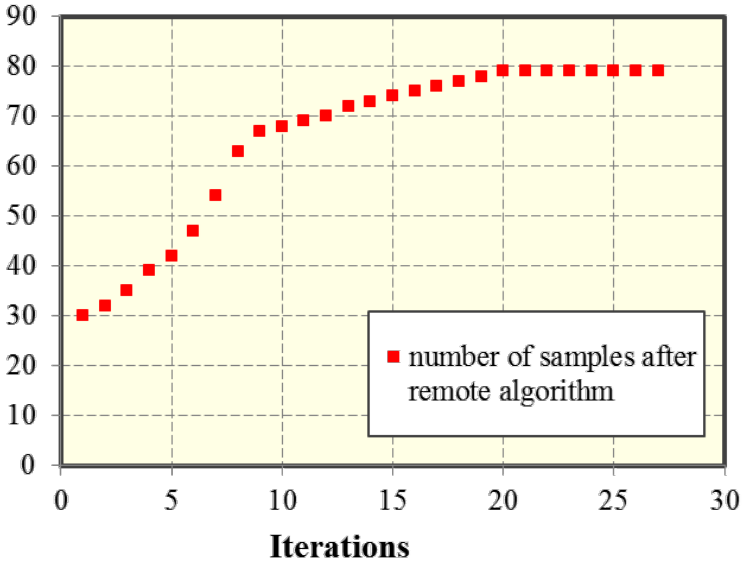
1. More trials on a potentially good match
2. Keep exploring the unknown space

The problem is that there is currently no universal theory that can tell us when to switch the stage. Thus, a governing equation is created to control switching process gradually.

An example of the governing equation is shown in Figure 3.5. For each iteration, if the value on the upper plot is large and the value on the lower plot is small, we have more efforts on exploring unknown combinations, vice versa.




(a) Number of samples after first cut off




(b) Number of samples after second cut off

Figure 3.5: Governing equation that controls the number of cut off after (a) first and (b) second sampling unite.

## 3.3 Interface for the Calculation

To empower the platform all kinds of calculation, two class of calculation interfaces are added. One is the built-in function for equations and algorithms; another is the interface to executables.

### 3.3.1 Interface of Executable Files

The interface of executable files is the key part for us to incorporate the physics in our workflow. The compiled executable files like simulator are usually used in the petroleum industry to predict physical behaviors, and the input files of them are usually complex. For models with uncertainties, we need to launch lots of simulations of different realizations, and the automation becomes necessary.

#### *3.3.1.1 XML and CSV*

Our platform provides a universal interface to executables by a template-to-files generator that can generate text files by batch. The current version of the file generator takes in a CSV file and a template file. The CSV file is like a spreadsheet and contains the scenarios chosen. Each column is one uncertain variable, and each row is one scenario to be generated. The template file is an XML that has designed tags that tell the generator what to with the content inside each tag. The current version of the generator can be used to generate any text file but will need some special treatments when the target file has XML tags inside the file.

In order to deal with text content that is influenced by multiple uncertain variables, four types of XML tags are designed respectively representing a constant part, the value of a variable, a string to be executed as a python script and return another string, and a string to be provided by internal calculation. The descriptions of the tags are listed in Table 3.1.

Table 3.1: XML tag to generate text files by batch

| Tag | Explanation | Example |
|---|---|---|
| **v** | The value of this uncertain variable | <v>hf</v> |
| **ev** | The value of a variable based on an external dictionary. (like the grid size) | <ev>lx</ev> |
| **ex** | Extra part that will not be affected by uncertain properties or variables in external value. Will be treated as plain text | </ex><ev>nx</ev><ex> |
| **py** | Python scripts: all the tags inside will be combined and treated as a python script, will Execute the python script and return the printed message. | <py><ex>print(int( </ex><ev>lz</ev> <ex>))</ex></py> |
| **str** | String: Convert the contents within tag into a string and directly output | <str><ex>dimens </ex><ev>nx</ev><ex> </ex><ev>ny</ev><ex> </ex></str> |
| **root** | Marking the begin and end of all the contents | - |

With this powerful module, we can incorporate executables easily into our workflow. Moreover, since all the modules are platform independent, we can use the executables in any type of operating system.

### *3.3.1.2 Simple Examples*

With the functions mentioned above, our platform can deal with the properties that are influenced by different uncertain property. One example is the effective fracture width and fracture conductivity. Fracture conductivity is the product of fracture width and

fracture permeability, when we use the concept of equivalent fracture width to accurate the simulation convergence rate, the porosity of fracture need to be adjusted accordingly, and the grid permeability shall be calculated based on the two properties.

$$C_{fracture} = D_{equivalent} * K_{fracture}, \quad (3.2)$$

where $C_{fracture}$ is the conductivity of the fracture, $D_{equivalent}$ is the equivalent width of the fracture, $K_{fracture}$ is the permeability of the grid of the fracture.

An example of the input files is shown in the figures below. A template for the EDFM preprocessor is presented in the text editor in Figure 3.6, and contents after the 13th line are omitted. The structure is a standard XML with the designed tag. Three different realizations of this template are generated in Figure 3.7 according to the scenarios we designed.

```
1  <root>
2  <str><ex>
3  folder </ex><ev>filename</ev><ex>
4  ginclude CMG
5  simulator 3
6  grid  1
7  gridorigin 0.0 0.0 0.0
8  dimens
9  </ex><ev>nx</ev><ex> </ex><ev>ny</ev><ex>
   </ex></str><py><ex>print(int(</ex><ev>nz</ev><ex>+</ex><ev>n
   zup</ev><ex>))</ex></py><str><ex>
10
11 lx </ex><ev>lx</ev><ex>
12 ly </ex><ev>ly</ev><ex>
13 lz
   </ex></str><py><ex>print(int(</ex><ev>lz</ev><ex>+</ex><ev>l
   zup</ev><ex>))</ex></py><str><ex>
14 ...
15
16 </ex></str>
17 </root>
```

Figure 3.6: An example of the template for generating input of the EDFM preprocessor.

```
folder NONFWCA_nf_1
ginclude CMG
simulator 3
grid  1
gridorigin 0.0 0.0 0.0
dimens
132 132 4
lx 5280
ly 5280
lz 544
dx var
132*40.0 /
dy var
132*40.0 /
dz var
2*108.5 2*163.75 /
```

(a) The output of the first scenario

```
folder NONFWCA_nf_2
ginclude CMG
simulator 3
grid  1
gridorigin 0.0 0.0 0.0
dimens
132 132 4
lx 5280
ly 5280
lz 544
dx var
132*40.0 /
dy var
132*40.0 /
dz var
2*108.5 2*163.75 /
```

(b) The output of the second scenario

```
folder NONFWCA_nf_0
ginclude CMG
simulator 3
grid  1
gridorigin 0.0 0.0 0.0
dimens
132 132 4
lx 5280
ly 5280
lz 544
dx var
132*40.0 /
dy var
132*40.0 /
dz var
2*108.5 2*163.75 /
```

(c) The output of the third scenario

Figure 3.7: The first, second and third outputs generated by the templates.

### *3.3.2 Build-in Calculations*

As a platform built in python, it is easy to add calculations according to the needs. As the powerful tools used in previous projects, we give three examples including net present value (NPV) calculations, clustering, and mean square error.

#### *3.3.2.1 NPV calculation*

NPV calculation is an example of the calculation that is based on equations the human belief of the environment. The equation we used here is based on the interest model:

$$\mathrm{NPV_n} = -\mathrm{C_i} + \sum_{j=1}^{n}\left[\left(\mathrm{I_j} - \mathrm{C_j}\right)\frac{1}{(1+\mathrm{R_{Discount}})^{\frac{j}{12}}}\right], \tag{3.3}$$

where $\mathrm{C_i}$ is the initial cost for drilling and completion, $\mathrm{R_{Discount}}$ is the discount rate that is used to convert the money to be earned from future to present, $\mathrm{I_j}$ is the income of the $j^{th}$ month, $\mathrm{C_j}$ is the cost for maintenance of the $j^{th}$ month, n is the total number of the month for production. The equation for $\mathrm{C_i}$, $\mathrm{I_j}$ and $\mathrm{C_j}$ may be different according to the problem and a simplified version is provided below:

$$\mathrm{C_i} = \mathrm{C_{well}} \times \mathrm{N_{well}}, \tag{3.4}$$

$$\mathrm{I_j} = \mathrm{P_{oil}} \times \mathrm{V_{oil,j}} + \mathrm{P_{gas}} \times \mathrm{V_{gas,j}}, \tag{3.5}$$

$$\mathrm{C_j} = \left(\mathrm{C_{fixed}} \times \mathrm{N_{well}} + \mathrm{C_{water}} \times \mathrm{V}_{water,\mathrm{j}}\right) + \Big(\mathrm{P_{oil}} \times \mathrm{V_{oil,j}} \times \left(\mathrm{T_{oil}} + \mathrm{T_{ex}}\right) + \mathrm{P_{gas}} \times \mathrm{V_{gas,j}} \times \left(\mathrm{T_{gas}} + \mathrm{T_{ex}}\right)\Big), \tag{3.6}$$

where $\mathrm{C_{well}}$ is the cost of a well, $\mathrm{N_{well}}$ is the number of wells in the model; $\mathrm{V_{oil,j}}$, $\mathrm{V_{gas,j}}$, and $\mathrm{V_{water,j}}$ are the oil, gas, and water production at $i^{th}$ month respectively; $\mathrm{P_{oil}}$ is the price of oil, $\mathrm{P_{gas}}$ is the price of gas; $\mathrm{C_{water}}$ is the cost for water disposal, $\mathrm{C_{fixed}}$ is the cost to maintain a well in a month, $\mathrm{T_{oil}}$ is the tax rate of oil, $\mathrm{T_{gas}}$ is the tax rate of gas, and $\mathrm{T_{ex}}$ is the sum of other taxes.

#### *3.3.1.2 Clustering*

Clustering is an example when we are directly utilizing packages from the community. Knowing all the details about a single algorithm may be helpful for one's knowledge of the problem, yet there is no necessity to reinvent the whales. As a language with tons of well-optimized packages, we can fully explore the accomplishments of human knowledge.

With the clustering results from Skitlearn, we can easily cluster the high dimensional data and get the center of each cluster, and thus save time for what we care for.

#### *3.3.1.3 Modified Mean Square Error*

In practice, we use mean square error to evaluate the deviation of the simulation from real production. The equation of mean square error is listed.

## 3.4 STATISTICAL MODELS

The statistical models are used to approximate the reality of unknown in most of our workflow. The benefits and limitation of the models we used in the previous projects are listed below.

### 3.4.1 Polynomial

Polynomial is a kind of linear regression that can be used to predict the model with multiple inputs. There are well-developed regularization methods to deal with overfitting.

### 3.4.2 K-Nearest Neighbor

K-Nearest Neighbor predicts base on the data near the region, the value it predicts is a weighted value of nearby points and thus is not able to give a value that is larger than the original dataset. Also, if there are many points in the dataset, comparing the distance between points will have a noticeable cost.

### 3.4.3 Artificial Neural Network

The artificial neural network is a prevalent model that has the potential for very nonlinear problems. There are lots of tools and methods to improve the performance of ANN, yet applying it to a specific problem may not be easy. The detailed introduction of the ANN implementation is provided in the next chapter. ANN has more computation cost while training, but it can make a large batch prediction in no time. The detailed implementation of ANN is to be discussed in Chapter 4.

### 3.4.4 Data Transformation

The uncertain properties have different units and range, which will cause trouble sampling and applying statistical models. For the simplicity of the workflow, all the uncertain variables are converted into the same scale by some kind of transformation.

The target range of the transformation in this study is from 0 to 1. Moreover, the transformation is done by linear stretching the maximum and minimum value to 1 and 0. Certainly, there is a property that influences the model exponentially, as the permeability. Moreover, there is a property that the maximum value can be several times that of the minimum value. Another type of transformation may be better for some specific case, yet investigations are needed to prove that.

Another thing worth mention is about the transformation of the response variable. We can design a different way of transformation according to the problem we are working on and the statistical model we are using. Take the ANN used in HM for example. ANN is good at telling if the two inputs have a significant difference; when the values are very close, the prediction may not be as well. For history matching, we want to explore the scenarios that result in low objective function value. With the transformation, the relatively low values can become large, and the difference can be distinguished better. Some of the transformations are listed below in Table 3.2.

Table 3.2: Type of data transformation methods that are frequently used

| Transformation | Equation | Description |
|---|---|---|
| **Linear** | $Y_n = \frac{Y_{old} - Y_{min}}{Y_{max} - Y_{min}}$ | $[Y_{min}\ , Y_{max}]$ is the range of Y. |
| **Log** | $Y_n = log(Y_{old})$ | |
| **Reverse Log** | $Y_n = \frac{1}{Y_{old} + E}$ | E is a small value to prevent zero denominator |
| **Square reverse** | $Y_n = \left(\frac{1}{Y_{old} + E}\right)^m$ | m is the exponent that can be assigned according to the need |

## 3.5 Visualization of Higher Dimensional Data and Model

Data with higher dimension usually contains more information, and that information is usually easier to understand when appropriately plotted. However, for data more than 3D, design the right way to visualize is not easy. In this section, we introduce a few ways to visualize and study the higher dimensional data and model on it.

### 3.5.1 Parallel Coordinates Plot

Parallel coordinates plot is a common way to visualize and analyze high-dimensional data. For an n-dimensional space, n evenly spaced parallel vertical lines are drawn. Each point in this space is represented as a polyline with vertices on each axes representing the value at a different dimension. An example of the parallel coordinates plot is shown in Figure 3.8. A detailed discussion of the interactive version of this plot is to be demonstrated in chapter 7 with an actual field case.

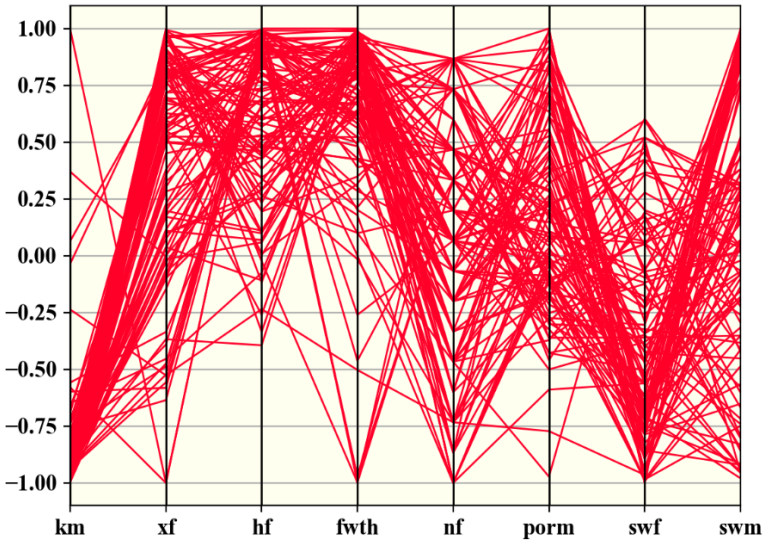



Figure 3.8: An example of the parallel coordinates plot.

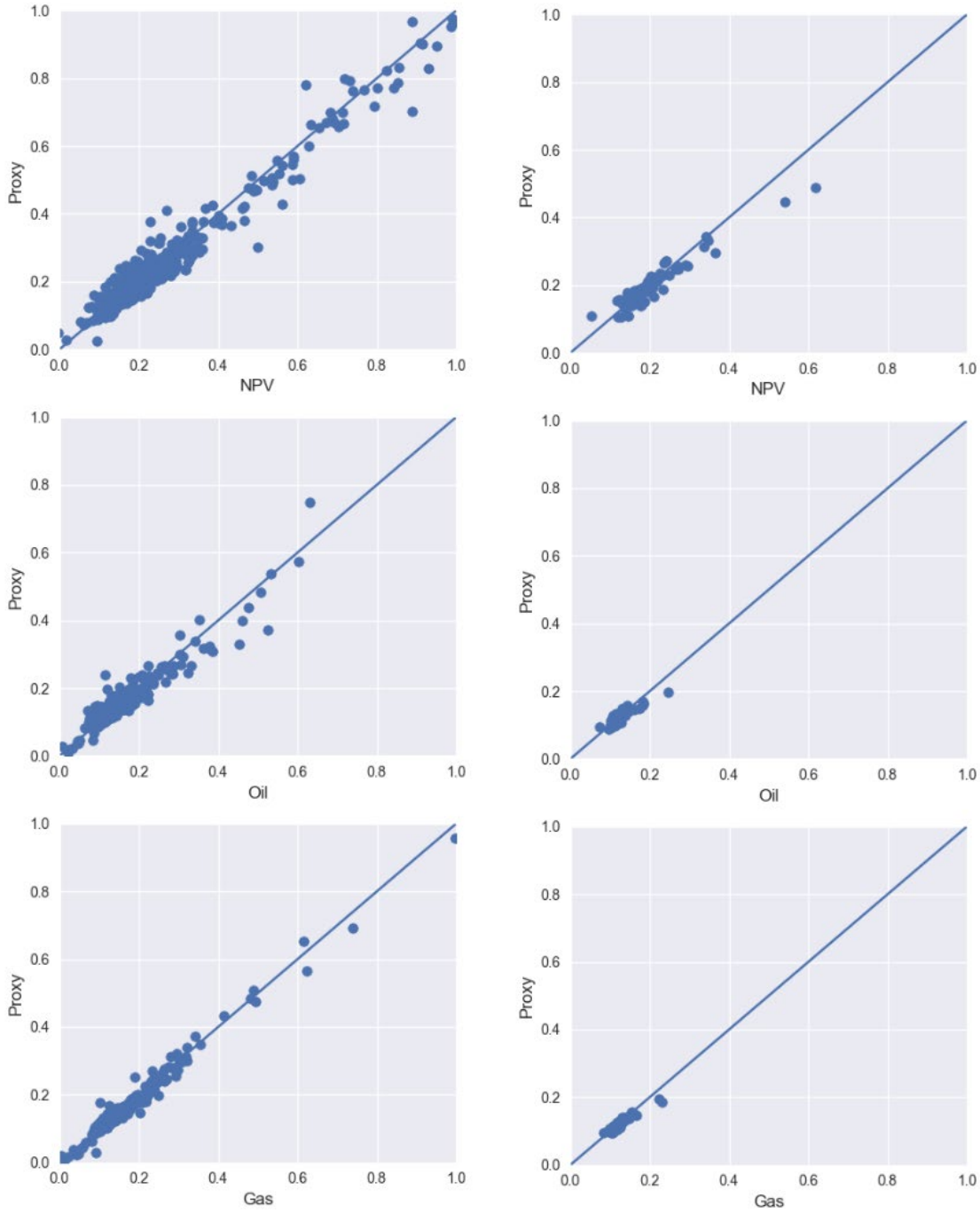
### 3.5.2 Model Diagnose

The model of high dimensional is usually considered a black box, and thus it is essential to understand whether the model is good enough for the problem.

#### *3.5.2.1 Compare model prediction and calculation results*

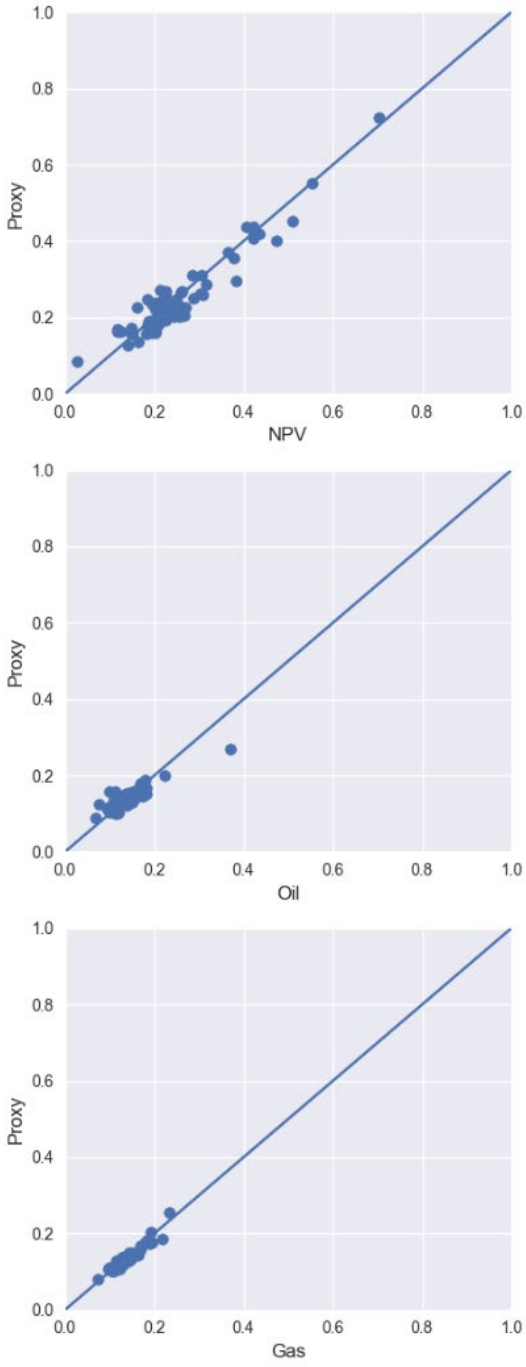
The most intuitive way to diagnose the model is directly comparing the proxy prediction and the original dataset. Usually, we do the cross-validation by dividing the original dataset into training and testing dataset. The training set is used to build the model; the testing set is used to check if the model can be generalized to the unknown dataset. For models that involve training, how much time to train may affect the performance of the model. Thus the dataset is divided into three groups; the developing dataset is added to select the model. Figure 3.9 shows the model's performance in training, developing, and testing dataset. The values are normalized to stay in the range of 0 to 1. If most points are

close to the line $Y = X$, then the model performs well on this dataset. If the model performs well on all the dataset, then the model has no obvious problem and thus can be used in the study.




(a) Training dataset (b) Developing dataset

Figure 3.9: Model performance on (a) training, (b) developing and (c) testing dataset.




(c) Testing dataset

Figure 3.9 continued

#### *3.5.2.2 Relative error for multiple responses.*

For a model that has multiple responses, it is hard to have a view on the overall performance. With Figure 3.10 we can plot the relative error into a heat map. For the plot shown, the dark color represents low error while the light color represents a high error.

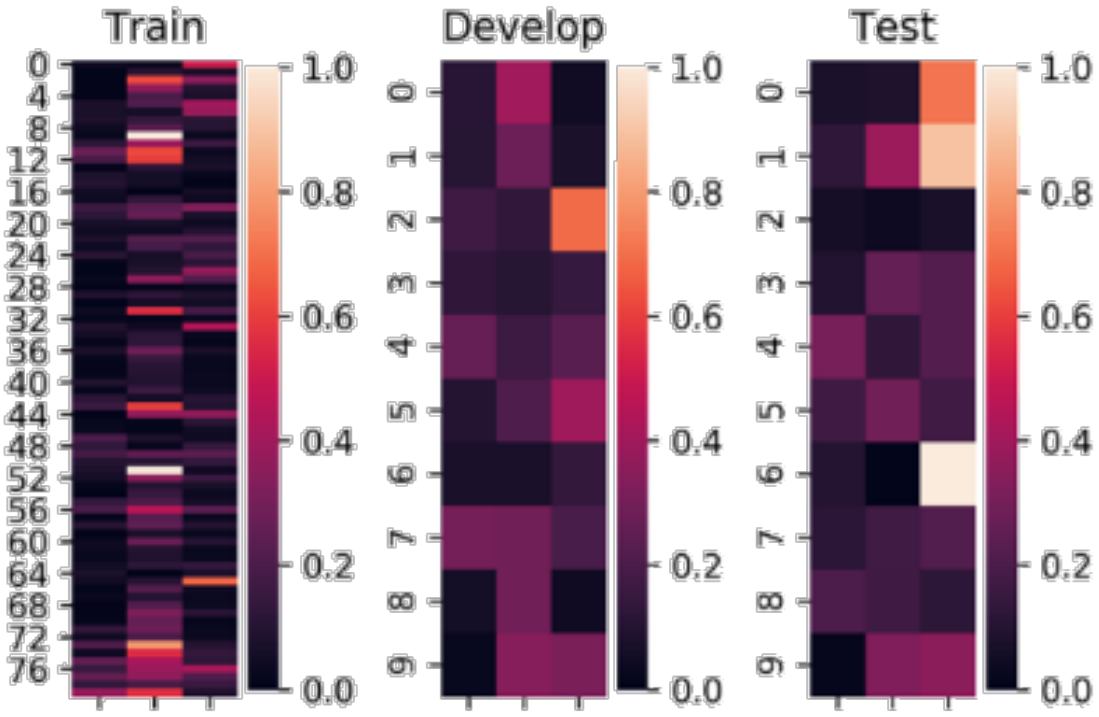



Figure 3.10: Relative error heat map for different datasets.

### 3.6 WORKFLOW EXAMPLES

Each module of the platform can be used directly for a purpose, but they are more potent while used together.

#### 3.6.1 History Matching

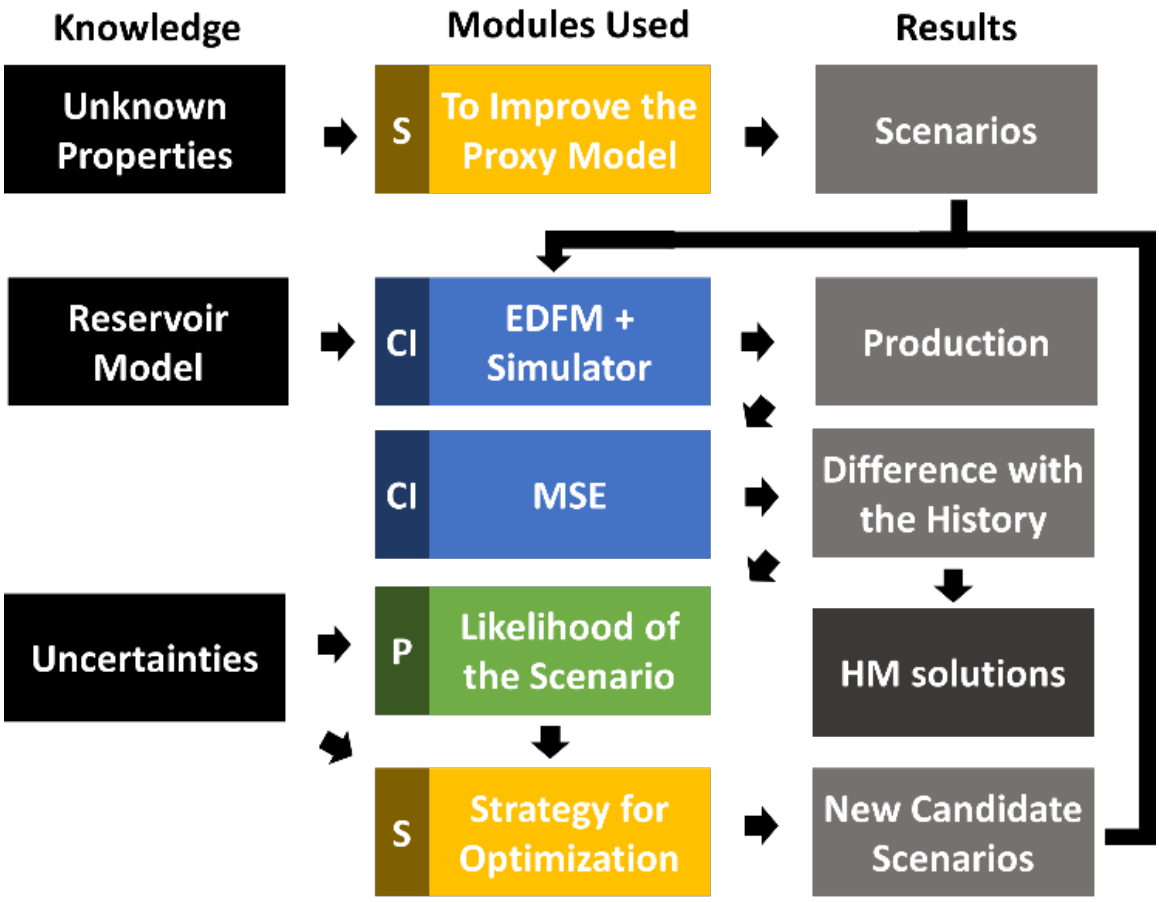



Figure 3.11: Generalized history matching workflow.

An example workflow of history matching using iterative response surface method is shown in Figure 3.11. The field case using this workflow will be discussed in Chapter 5 and Chapter 6.

### 3.6.2 Well Spacing Optimization

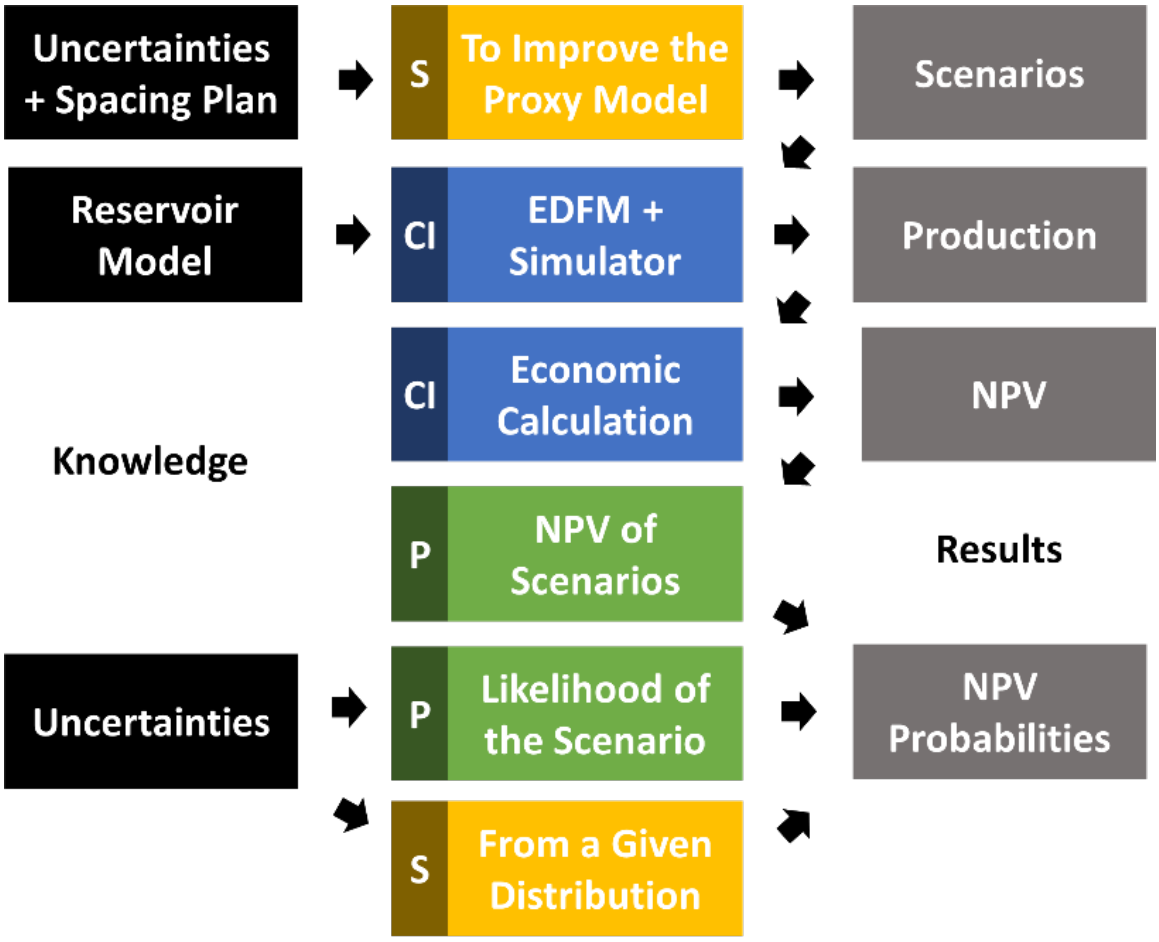



Figure 3.12: Generalized well spacing optimization workflow.

An example workflow of well spacing optimization is shown in Figure 3.12; the field case using this workflow will be discussed in Chapter 6.

# Chapter 4: Neural Network as a Predictive Model

This chapter is about the practice to use the artificial neural network (ANN) as a predictive model for our platform. After a brief introduction to the basic knowledge behind the ANN, the implementations to preprocess data, deal with the overfitting, and solve the limitation of datasets available are presented.

## 4.1 Brief introduction to ANN

The ANN is an old tool that enabled new power in various fields after the progress in recent years. Its family includes the convolution neural network (CNN) and the recurrent neural network (RNN) that can be applied to complex problems like computer vision and natural language processing. ANN can also be used merely as a regression model to capture the changes in nonlinear behaviors even when we do not have much knowledge about the problem.

One crucial feature of ANN that we are going to take advantage of is that it can be more flexible than most other predictive models and thus can fit nonlinear response surface with better overall performance. However, in the context of history matching, we need to deal with the problems like insufficient training set, choice of the ANN structure, and overfitting.

For a regression model with constant numbers of input and output variables, a basic ANN is enough. The first crucial feature of ANN that we are going to take advantage of is its flexibility. Another benefit is that we can predict multiple objects at the same time using any decided number of input variables.

Building a typical ANN needs to decide the number of layers, the number of neurons in each layer, how neural are connected, the activation function, and the optimization method. There are also tools to improve performance like early-stopping, initialization methods, and dropout layers. Since this is not a paper in machine learning, the main component of the ANN selected for the platform is introduced in Table 4.1 without diving into the detail of the technics. The performance of ANN is affected by its structure and the complexity of the problem. A good model will need iteration on the

hyperparameters that control its structure. The default structure in the platform is merely the starting point for the process and is described in Table 4.2.

Table 4.1: The methods of choice in our platform

| **Methods** | **Implementation** |
|---|---|
| **Library Used** | Keras with Tensorflow as the backend |
| **Model Type** | Sequential model |
| **Optimizer** | SGD with default settings |
| **Activation Functions** | Relu for all the layers and neural |
| **Layer Types** | Fully connected layers for all the hidden layers |
| **Number of layers** | 3 and above |

Table 4.2: The structure of ANN

| **Layers** | **Number of Neural on the Layer** |
|---|---|
| **Input Layer** | Number of uncertain property (Nu) |
| **Hidden Layer 1** | Nu *2 |
| **Other Hidden Layers** | Nu + No<br>(number of layers to be added depends on the demand) |
| **Hidden Layer N** | No*2 |
| **Output Layer** | Number of uncertain property (No) |

### 4.2 PREPROCESS THE DATA

Any type of data can be used as the inputs of ANN, yet proper editing of the dataset can make the implementation easier. For our workflow, all the variables are converted to float numbers so that there is no need to do the special treatment inside the ANN. The scale and units of the data are taken care of by normalizations in the sampling module.

Like a typical way to train and access the statistical model, the dataset will be divided into three parts, the training dataset, the developing dataset, and the testing dataset. As implied by the name, the training dataset is used to train the ANN. The developing dataset is used to evaluate the model during the training process and help us select the model. The testing dataset is regarded as unknown during the training process and used to evaluate the model in case we overfit both the training and developing datasets.

### 4.3 MODEL SELECTION

The performance of ANN can perform badly unless adequately tuned. Overfitting is the biggest issue when we use flexible statistical models. It is vital for us to monitor performance during the training process. Figure 4.1 shows us a single ANN's performance on training, developing, and testing dataset thought the training process. For this example, 200 epochs of training are performed in each iteration. At the end of the $30^{th}$ iteration, the model went through 6000 epochs of training. At the end of each iteration, the model is used on the datasets, and the response is compared with the real ones. MSE is calculated, and the average of all the responses are plotted in the figure. The diagnose plots mentioned in section 3.6.2 are also automatically generated for each dataset. The zoomed plot of diagnosing plots after $1^{st}$, $10^{th}$, and $30^{th}$ iterations are shown in Figure 4.2. The $10^{th}$ model is regarded as the best model considering the overfitting as it is the turning point for developing and test set in Figure 4.1. The $1^{st}$ model is underfitting as the points are not following the x = y in the scatter plot. The $30^{th}$ model has the best performance on the training dataset, yet its worse performance on developing set indicates a bad ability to extend the prediction on the unknown dataset, which is proved by the testing set. Usually, the performance on testing and developing will be similar; the obvious discrepancy of those two is due to the randomness caused by insufficient dataset and lack of redundancy.

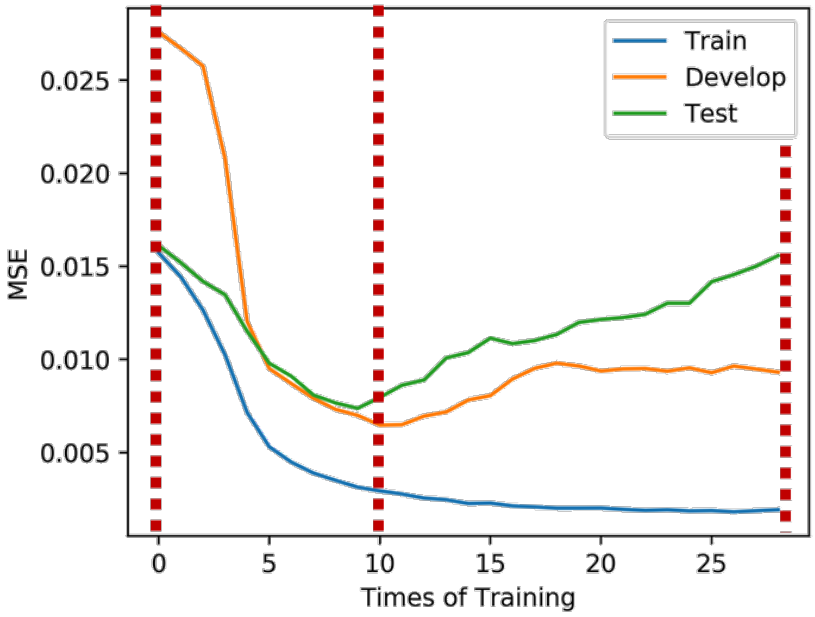
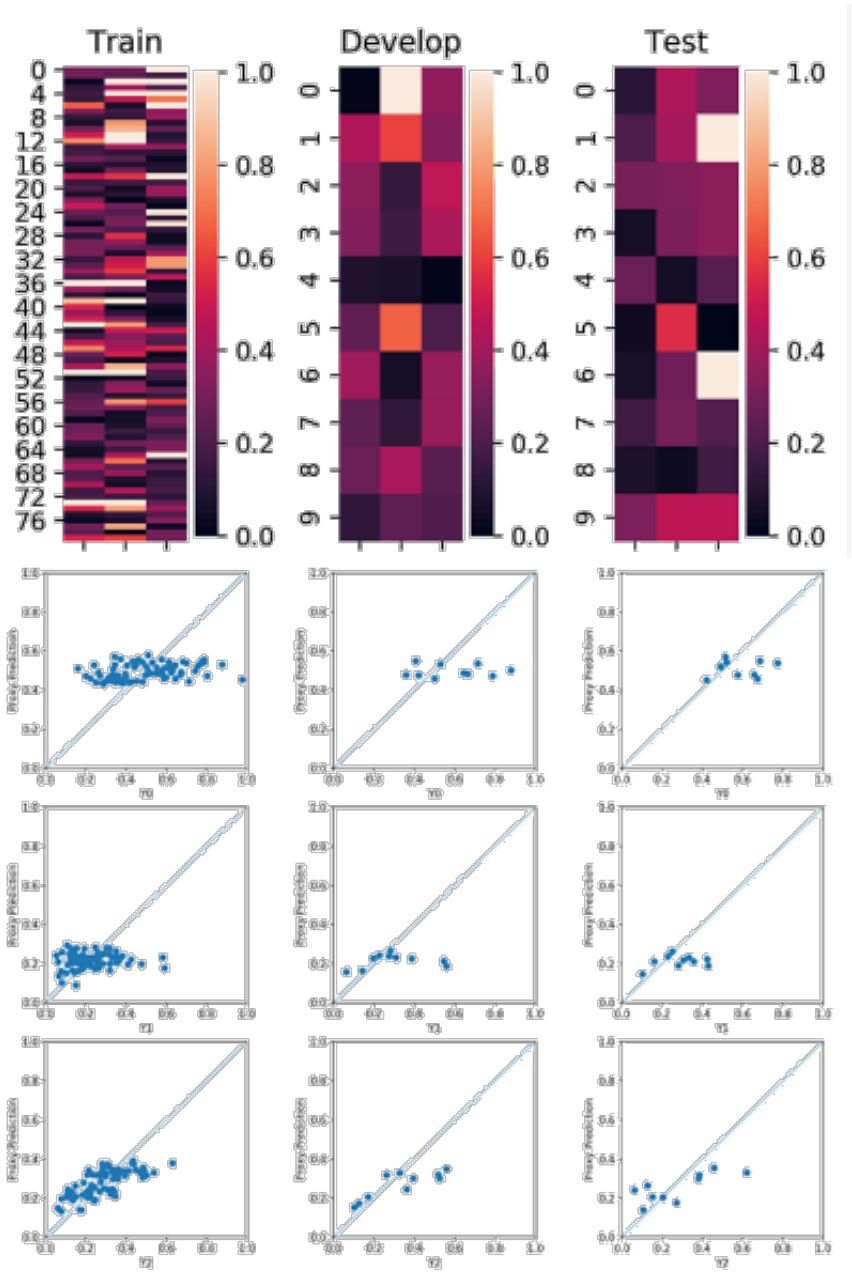
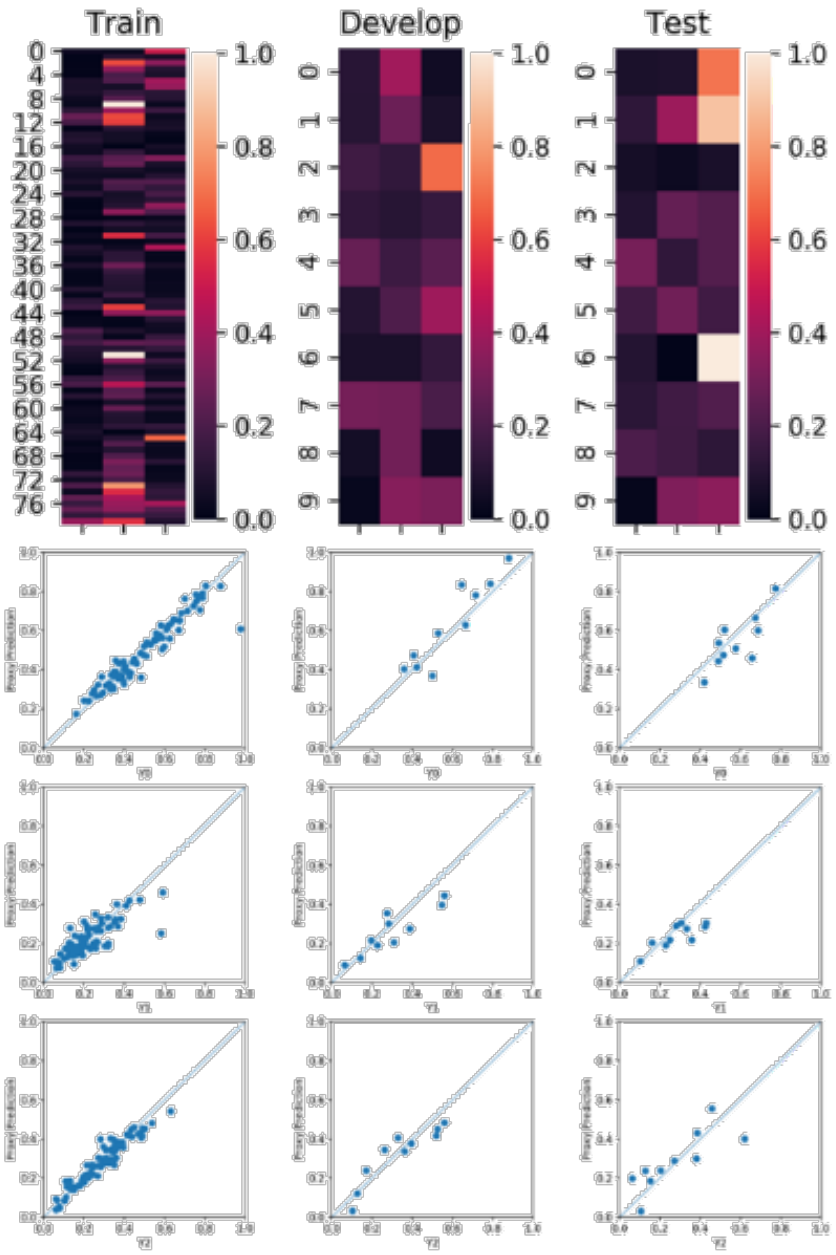



Figure 4.1: Model performance of MSE on a different dataset with different times of training.



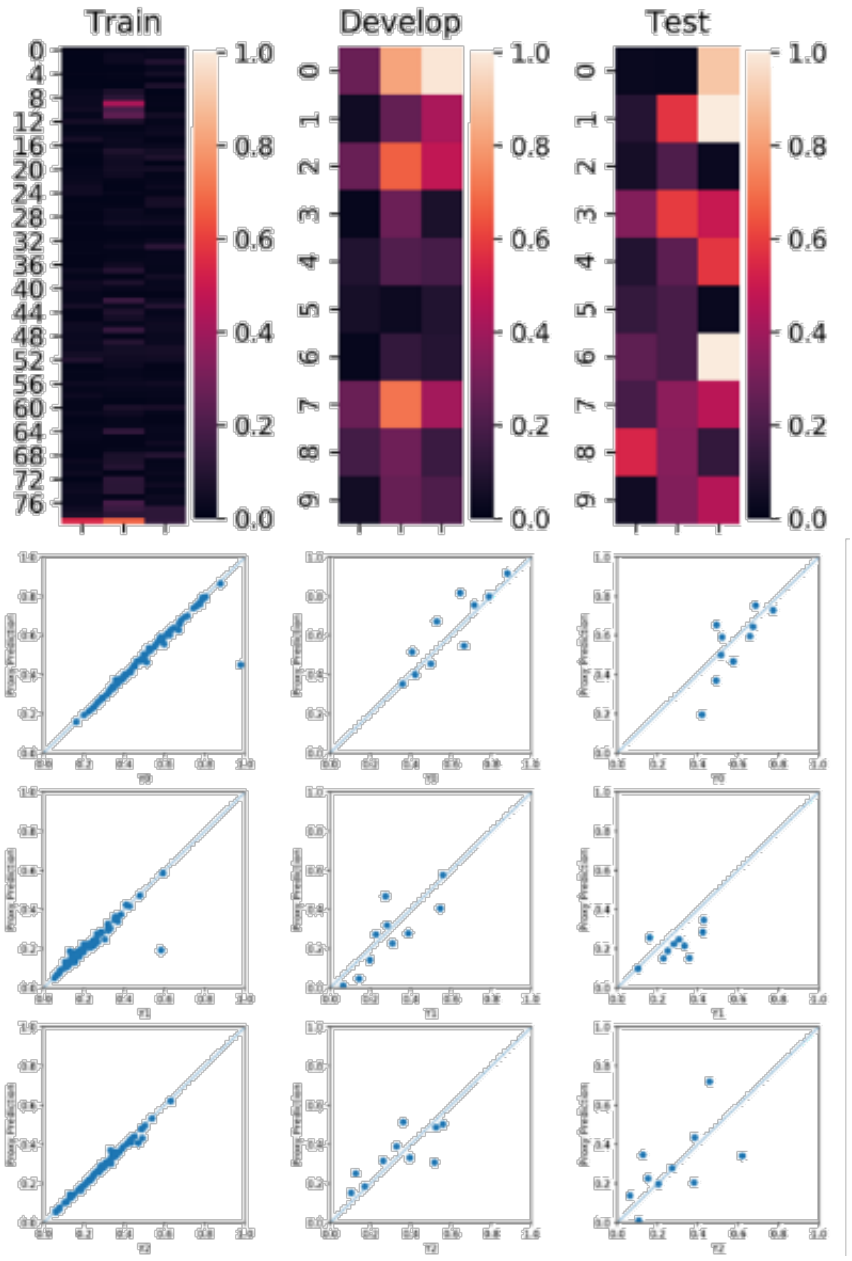

(a) Model performance after the first batch of training

Figure 4.2: Model performance when finishing (a) 1, (b) 11, and (c) 30 batches of training.




(b) Model performance after the 11th batch of training

Figure 4.2 continued




(c) Model's performance after the 30th batch of training

Figure 4.2 continued

There are other criteria to judge the usability of the model that can be used depending on the problem we are facing. The example of them includes, (1) direct

comparison of proxy prediction and real value, (2) overall RMSE of different model, (3) R-square. The platform can add the wanted feature easily.

### 4.4 TYPICAL CHALLENGES

In our platform, ANN is used as a proxy model, and there are two primary usages. The first one is to use the model for prediction directly. The other is used in the workflow to reduce the number of simulation needed. The problem we are facing is different.

For history matching problem, the first concern for using ANN is that we usually do not want to run too many simulations, and thus have a considerably small training data set. For our iterative response surface workflow, we are adding batches of data as we choose new scenarios to be simulated, which lead to even fewer data when we start the workflow with only the number of sensitivity runs.

ANN and other deep learning models are initially designed for more complex problems and thus need typically millions of elements in the dataset to boost the performance. The dataset is usually divided into the training, development, and testing sets or even more for different purposes. To take advantage of the precious dataset that we can obtain, only training and development sets are distinguished and used with several strategies to be introduced in the application part.

Simulation results, which are used to train the ANN, are also different from real-world data since they are usually with some assumption that excludes some less critical factor and have less measurement accuracy issue. When we are not working on a problem with very high non-linearity, the response surface is considered smooth on a small scale.

With more layers and more neural units in each layer, the ANN can be adjusted to fit more complex response surface, and the cost of a deeper and larger ANN will be the computation cost. Luckily, time spends on training ANN with our limited dataset will be ignorable compared with the reservoir simulation. Thus, a considerably large ANN can be used if necessary.

An overfitting structure is chosen for our workflow considering the unknown knowledge and our purpose. For a certain problem, ANN can be adjusted from underfitting

to overfitting by adding layers and neural units. There could be an optimum structure for this specific problem, but we do not know how nonlinear the problem is beforehand. Finding an optimum structure will need a large dataset and iterate through the hyperparameters of the ANN. The optimum structure is also changing as more points are added to the training dataset. Also, there are lots of well-developed strategies to reduce overfitting.

It seems a luxury to divide a testing dataset to do the cross-validation. However, we are also not expecting a perfect proxy model that can replace the simulator. An idea proxy model for our history matching workflow will include two features. First, an acceptable prediction is given if the matching error is high or low. Second, for the cases with low matching error, we want them to be predicted more accurately than the ones with a higher matching error. As mentioned later in the sampling strategy part, we can take advantage of the overfitting.

When we directly use the model for prediction, the primary issue is overfitting. One of the examples is to use the properties and productions of wells from one area to predict the production of a candidate well. For this kind of problem, there may be confliction between the data in the training dataset. It is caused by the fact that the input features cannot adequately represent the difference between wells. Some of the properties like the porosity and permeability are the statistical summary of the matrix around the well, and thus relevant information may be lost. We can add more detailed features to reduce this kind of problem yet gathering the data may not be physically and economically possible. When the model tries to fit two values that have very similar input features but distinct output, it may cause extreme values around those values.

When we want to reduce the number of simulation with the proxy model, the problem is insufficient training dataset. It is also a more common situation to be discussed in this study. Unlike the dataset gathered from reality, we can assume that the prediction of the model is continuous as the features change continuously. This assumption may be compromised when geomechanics or complex phase behaviors are introduced. It is still sound at regions away from those critical situations. Although we have more tolerance to

overfitting, the new problem occurs since we need to balance the accuracy of the model and computational cost of simulation.

Although the cost of running simulation is smaller than gathering data from the real world, it is still not plausible to run too many simulations. Moreover, as described in the previous chapter, more scenarios are at the place of interest and less are at the unlikely zone. If we divide the dataset into three part like usual, it is possible we may lose important information.

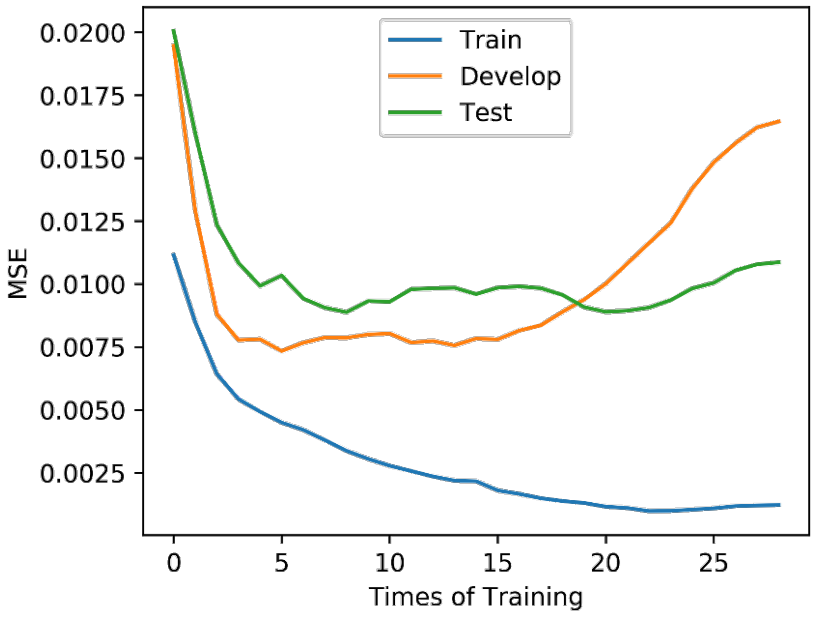



Figure 4.3: Performance of another ANN when the only difference is the random seed.

Another issue problem of using ANN is that we have stochastic processes during the optimization of the gradient. We don't know exactly how well a model will be trained even with exact same dataset and ANN structure. Figure 4.3 shows another model trained base on the same setting as that of the model showed in Figure 4.1 but have a different random seed. Also, sometime the model will stop improving when it is not good enough.

### 4.5 IMPLEMENTATION

We choose bagging to overcome the overfitting problem caused by redundancies in the structure of ANN.

Bagging is mostly used in the random forest to solve the complex problem with multiple not-so-complex models. Results from multiple models are gathered together and generate the final result by a decision rule. Bagging on ANN is usually replaced by a strategy called drop out which deactivate several neural while training for each batch. The dropout layers are not used here in order to build models with all the data from the dataset.

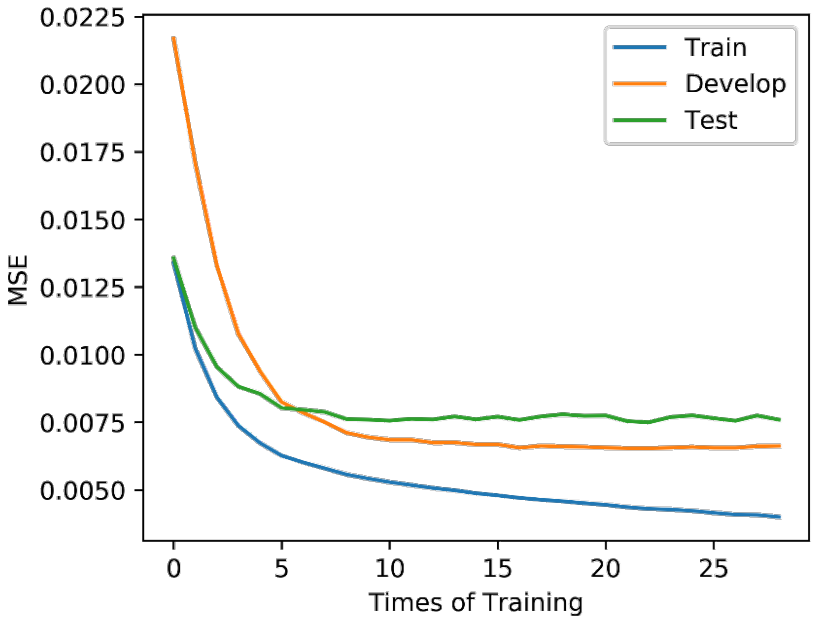



Figure 4.4: Performance of a bagged model.

As mentioned in the training set section, our dataset is usually divided into parts to reduce overfitting. When bagging is used, multiple ANN will be trained, and thus we can divide the original dataset into different pairs of training set and development set so that every point from original dataset are used to train several ANNs. After training, the lost function on the training and development set is recorded for later prediction. The final prediction is a weighted average of all the selected models.

The recommended workflow to bag the ANN requires us to store the model at each iteration:

1. Decide the number of models to be used N1, the number of new models to be trained N2, the number of models to be stored N3(>N1), the number of models to be trained initially is N4(>N3).

2. Randomly divide the data we have into training and testing data set with the ratio 8:2 and train one model.

3. Evaluate the model with both the training set and the testing set, calculate a weighted average as the score. Store both the model and score to a paired list.

4. Repeat step 2 and step 3 until we get N4 models, save the top N3 models and delete the rest. The top N1 saved models and score will be used to dot the prediction together.

5. After we get new data or when we want to restart the training, set N4=N2+N3.

6. Train the models we stored. The training process is the same as that in step 2 and step3.

7. Train new models with new dataset until the request of step 4 is satisfied.

The reason not all stored models are used is that we want to rule out models that are not predicting well but training them to cost less than training a completely new model. However, we are still training new models in case the stored model goes wrong and do not improve.

When we make the prediction with N1 models selected, the result of bagged models, $Y_{final}$, can be regarded as a weighted average.

$$Y_{final} = \frac{\sum \frac{Yi}{Si}}{\sum (\frac{1}{Si})}, \tag{4.1}$$

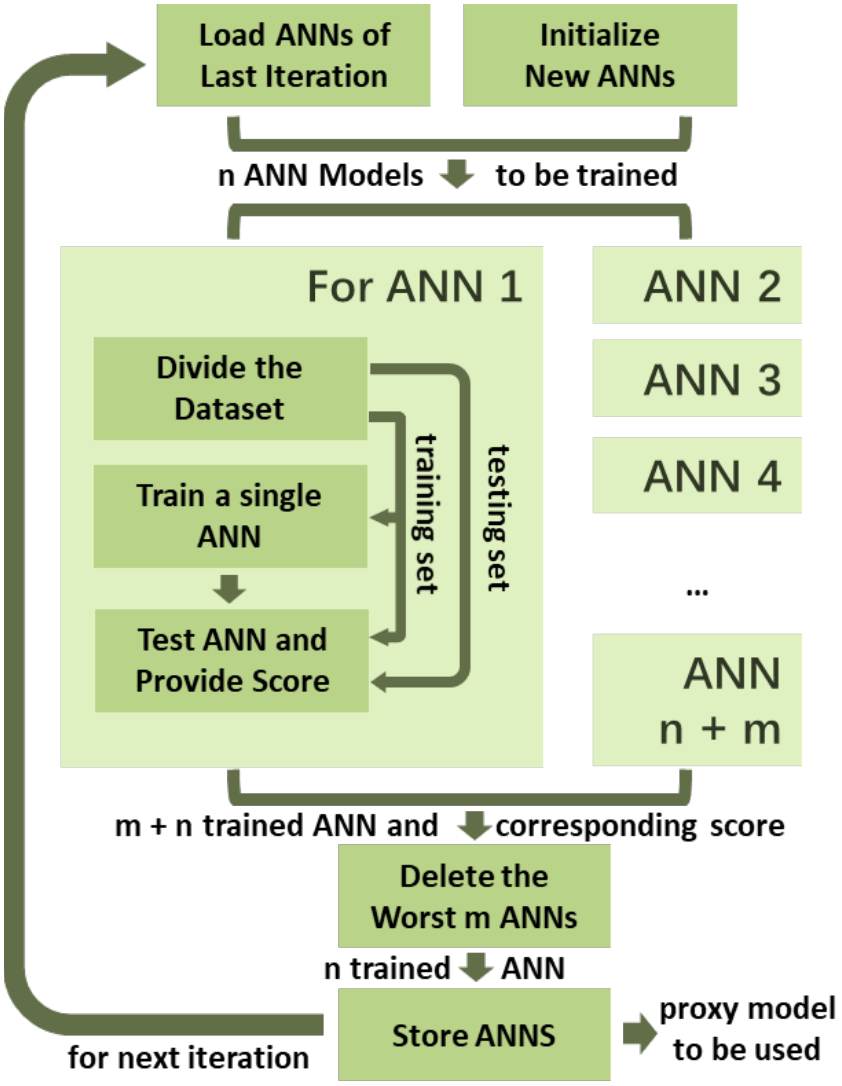



Figure 4.5: ANN training strategy for proxy-based history matching workflow.

We automated the entire workflow with Python and used Keras and Tensorflow to build the ANN. Three fully connect layers are used as a hidden layer, and Relu is chosen as the activation function. Adadelta is the optimizer selected to perform the backward propagation.

The ANN built here may not be the optimum on performance or efficiency, but it can already boost the workflow.

## 4.6 CONCLUSIONS

The ANN is a powerful tool that can be used as a predictive model in the workflow to manage the unconventional assets and thus is implemented in the platform. The important points while using the ANN in our workflows are summarized:

1. As a model that can be made almost infinitely flexible, the design of the model can affect the performance and efficiency. A model with more neutrals and layers can be used on the more non-linear problem, but the computation time to train the model will increase accordingly. For beginners, a set of setting is provided in this chapter as a reference. It can be a starting point to iterate for a different problem.
2. Overfitting is the biggest issue for ANN because it may compromise the ability to generalize the model to an unknown dataset of the same problem. The dataset is supposed to be divided into three parts to train, develop and test the model. Model selection is necessary after model diagnosing.
3. For our specific problems in unconventional field management, the data that can be used to build the model is biased and insufficient. Bagging is used in the platform to deal with overfitting and take advantage of more data.
4. To use the ANN in the automated workflow where data are added consequently, a suggested training scheme is provided.

# Chapter 5: Optimization of Well Spacing in a Shale Gas Reservoir

This chapter presents an example of well spacing optimization that utilizes our platform in a shale gas reservoir with a few assumptions. First of all, a multi-fractured horizontal well is history matched to quantify the uncertainties of completion performance and matrix properties. Then, the uncertainties quantified are used to evaluate the performance of candidate well spacing plans. Finally, we test the effect of natural fractures on economics and demonstrate how we utilize both simulation and statistical models.

## 5.1 Reservoir Model

The example in this chapter is based on an actual field case in Marcellus Shale gas reservoir produced by a multi-fractured horizontal well. The detailed discussions on how this well is modeled for history matching can be found in work by Wantawin et al. (2017a, 2017b). Here, we are only introducing the basic information necessary to reproduce the model.

The black-oil model is used with modified fluid properties to capture the gas desorption effect. The reservoir comprises two layers with different matrix porosity, and the horizontal well is placed at the center of the bottom layer of the formation. Sixteen hydraulic fracture stages are evenly distributed along the horizontal wellbore, and 64 effective hydraulic fractures are generated, as shown in Figure 5.1. The reservoir is a dry gas reservoir with pressure initially above the dew point. Other important non-variable properties are listed in Table 5.1 (Yu et al., 2014).

Table 5.1: Basic reservoir and fracture parameters used in the simulation model

| Parameter | Value | Unit |
|---|---|---|
| Model dimension (x × y × z) | 5,000 × 2,000 × 135 | feet |
| Number of grid blocks (x × y × z) | 500 × 41 × 2 | - |
| Initial reservoir pressure | 5,100 | psi |
| Reservoir temperature | 130 | $^{o}$F |
| Total compressibility | $3 \times 10^{-6}$ | $psi^{-1}$ |
| Upper layer thickness | 95 | ft |
| Bottom layer thickness | 40 | ft |
| Matrix porosity (upper layer) | 7.1% | - |
| Matrix porosity (bottom layer) | 14.2% | - |
| Horizontal well length | 3,921 | feet |
| Number of stages | 16 | - |
| Cluster spacing | 50 | feet |
| Total number of fractures | 64 | - |
| Fracture width | 0.01 | feet |
| Gas specific gravity | 0.58 | - |

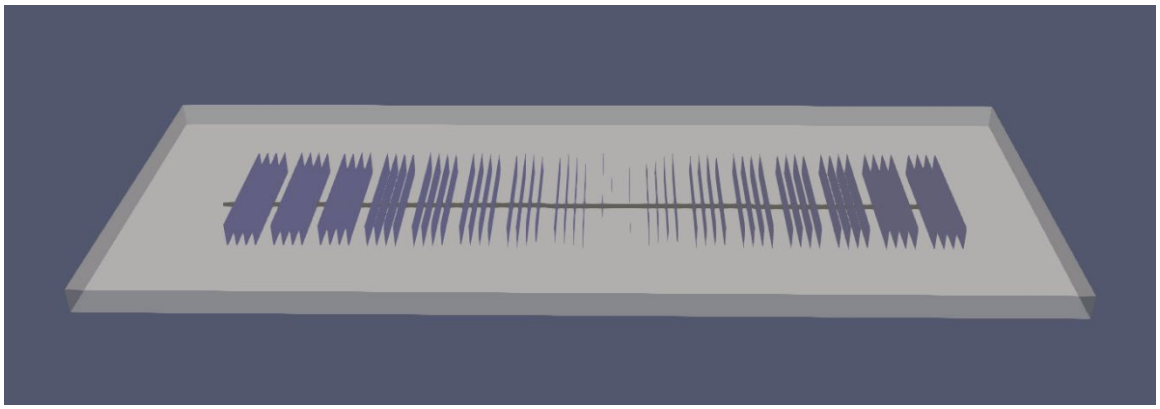


Figure 5.1: A basic reservoir model including 64 hydraulic fractures modeled using the EDFM method. The black line in the middle is the horizontal well; the blue surfaces vertical to the well are hydraulic fractures.

Uncertain parameters were selected from reservoir and fracture properties that have a large influence on the production and is hard to measure in reality. In this study, hydraulic fractures are assumed simple planner geometry and are described by fracture half-length, fracture conductivity, and fracture height. Fracture half-length measures the perpendicular distance from the wellbore to the fracture tip. Fracture conductivity is the product of

fracture permeability and fracture width. Fracture height measures the height of the hydraulic fractures extending from the bottom layer to the upper layer of the formation. The entire thickness of the bottom layer is believed fully penetrated. The uncertainty is, therefore, the extension of hydraulic fractures into the upper layer. Four uncertain parameters such as matrix permeability, fracture half-length, fracture conductivity, and fracture height and their effective ranges are summarized in Table 5.2.

Table 5.2: Four uncertain parameters and their effective ranges

| Uncertain Parameter | Unit | Low | High |
|---|---|---|---|
| Matrix permeability | nD | 100 | 1000 |
| Fracture half-length | ft | 300 | 500 |
| Fracture conductivity | mD-ft | 1 | 10 |
| Fracture height | ft | 40 | 135 |

The first 190-day production data after opening the well are used to calibrate uncertain parameters in the reservoir model during the automatic history matching process. For this practice, the bottom hole pressure (BHP) is used as the reservoir simulation constraint, and the gas production data is used as the history matching target.

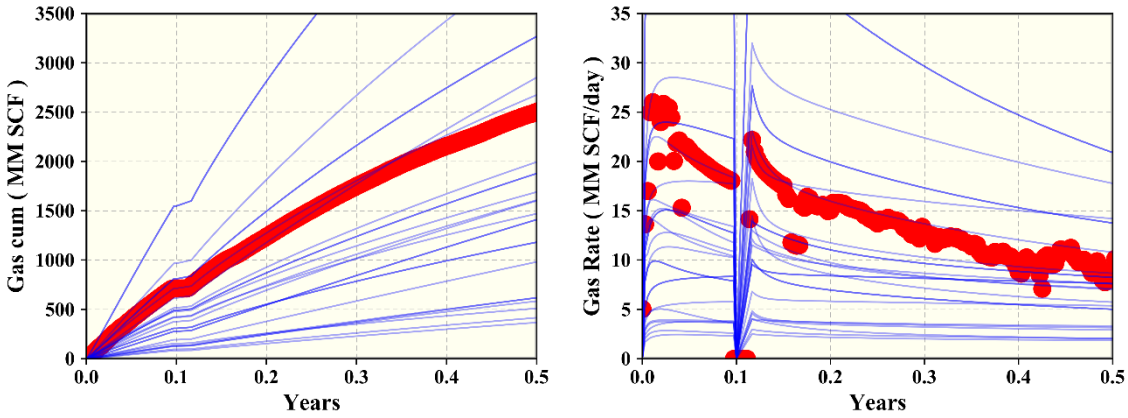
## 5.2 HISTORY MATCHING

The automatic history matching process is done with our workflow in section 3.6.1. This workflow is based on the iterative response surface method. Well-designed sampling strategy and proxy model are used in the automation which provides a more diversified solution set.

### 5.2.1 Initial Design

Twenty-five simulation runs were performed as the initial design based on the two-level full factorial design. The simulated production profiles are shown in Figure 5.2. It can be clearly seen that the actual production data shown in red dots are covered by the

simulation results represented by the blue lines, confirming that the history matching solutions can be found in the given range of uncertain parameters.




(a) Cumulative gas production (b) Gas flow rate

Figure. 5.2. Comparison of 25 simulation results with actual production data from the two-level full factorial design.

### 5.2.2 Progress and Results

The weighted root mean square error (RMSE) is used to quantify the discrepancies between the production history and the simulated results. The smaller the value of RMSE, the better match the simulated scenario is. A threshold value of 3000 is set to RMSE to filter the cases as history matching solutions in this case study. The value 3000 is chosen base on the visual estimation of the acceptable history matching error from the plot of production and simulation results.

During the workflow of iterative response surface method, 35 iterations are run, and for each iteration, 10 scenarios are evaluated. Among all the 375 simulations, 172 scenarios are regarded as the history-matching solutions. The acceptance ratio is about 46%. Figure 5.3 shows the weighted RMSE of the cumulative gas production during the history matching period; the points in red are regarded the history matching solutions. According to the plot, the weighted RMSE is gradually decreasing, thus suggesting better history matching solutions. The comparison of the simulation results of the 172 history

matching solutions with actual production data is plotted in Figure 5.4. As shown, reasonable matches for gas flow rate and cumulative gas production are achieved.

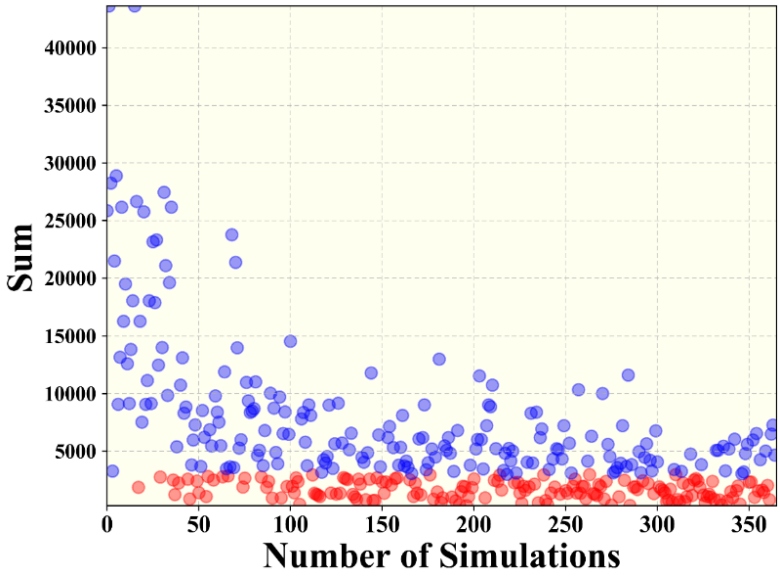
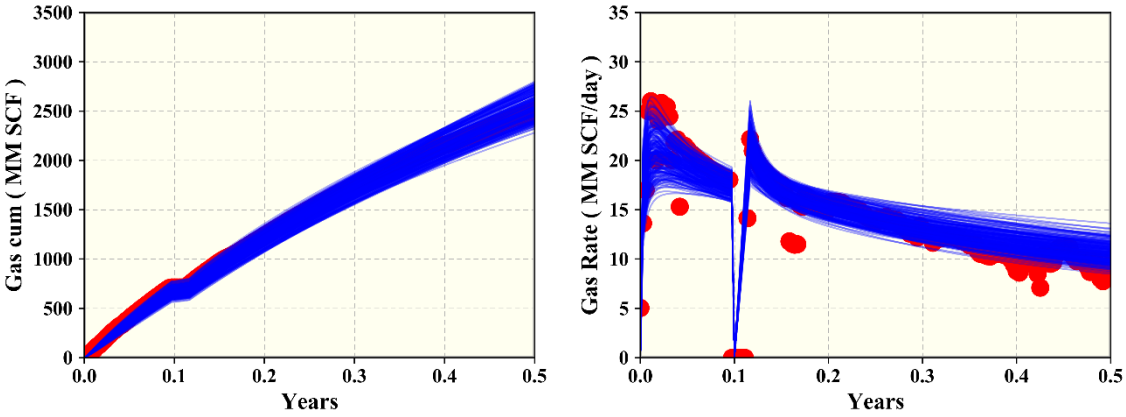


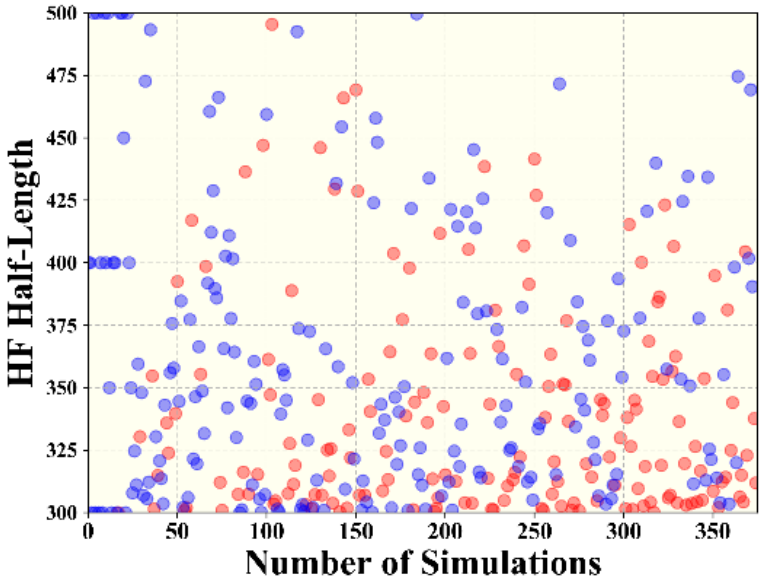

Figure 5.3: Weighted RMSE of all the simulation runs (red dots represent history-matching solutions, and blue dots are non history-matching solutions).




(a) Cumulative gas production (b) Gas flow rate

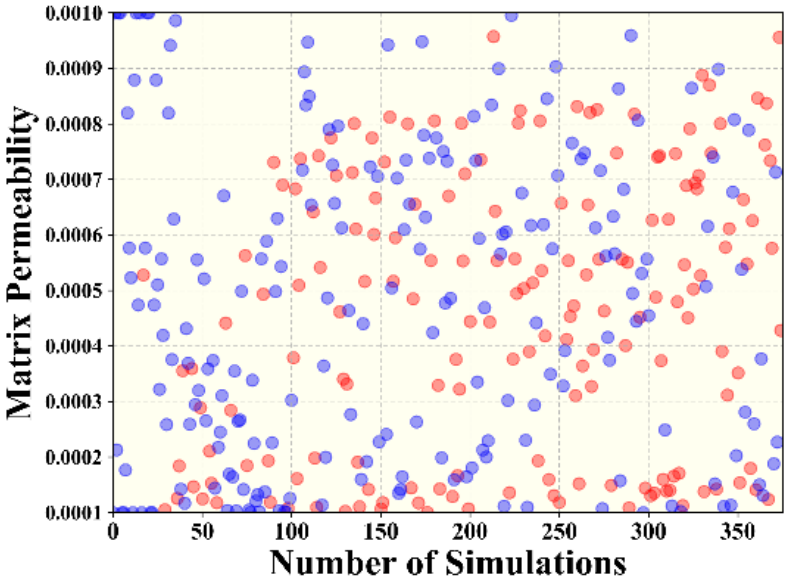
Figure. 5.4. Comparison of 172 simulation results of the history matching solutions with actual production data.

The reason we can still see badly-matched scenarios in the later iterations is that our algorithm is trying to explore the uncertain space thought the time. Figure 5.5 shows the progression of four uncertain parameters during the history matching process. As we can see, the scenarios chosen are trying to cover the uncertain space thought the time. This kind of behavior can prevent us from trapping into a single solution region that is found in the early stage. In this case, the value of matrix permeability can be used as an example. In the beginning, the scenarios with large matrix permeability tend to have large matching errors, and in turn, most algorithms will concentrate on the area that has a relatively low matching error. Our algorithm did not give up on this area, although it is also testing more scenarios with low permeability, several trials on higher permeability are still done due to our diversity control mechanism for each iteration. This mechanism is realized by the sampling strategy mentioned in section 3.2.2.

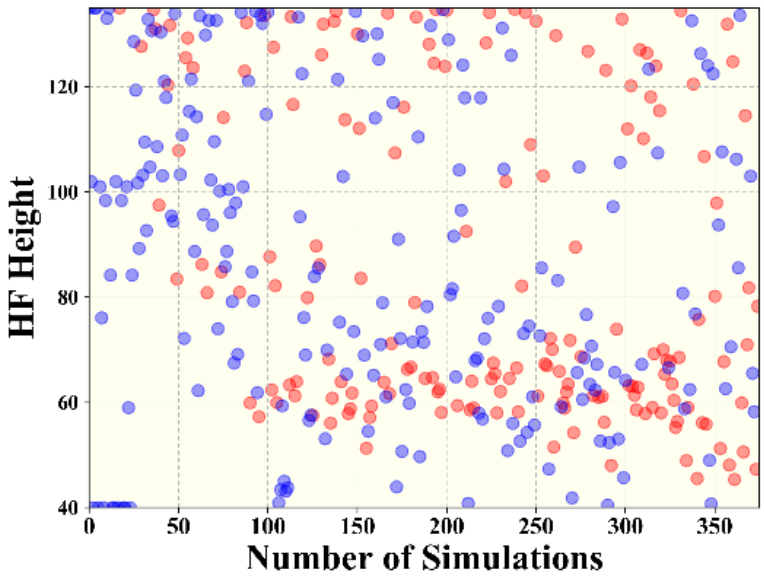



(a) Half-length of the hydraulic fracture

Figure 5.5: Progression of four uncertain parameters sampling during the history matching process (red dots represent history-matching solutions, and blue dots are non history-matching solutions).

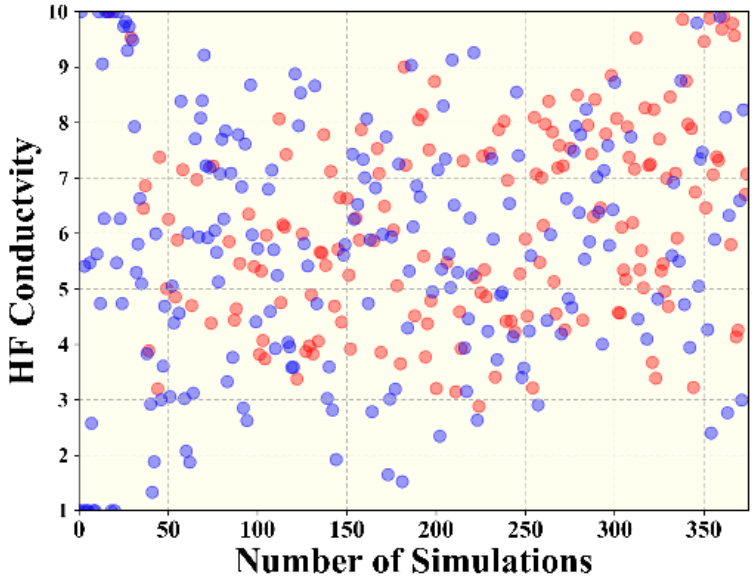



(b) Permeability of the reservoir matrix




(c) Height of the hydraulic fracture

Figure 5.5 continued




(d) The conductivity of the hydraulic fracture

Figure 5.5 continued

## 5.3 Well Spacing Optimization

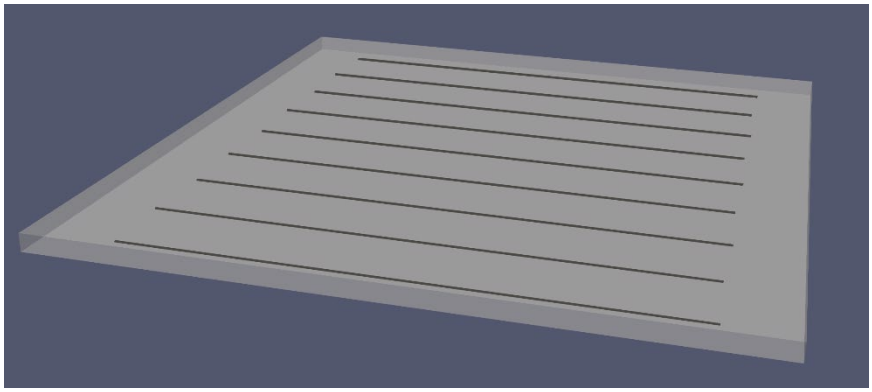
The optimal well spacing is crucial for the economic development of the unconventional reservoirs. Here we present an example that tests different well spacing plans based on the uncertainties evaluated by history matching.

### 5.3.1 The Spacing Plans

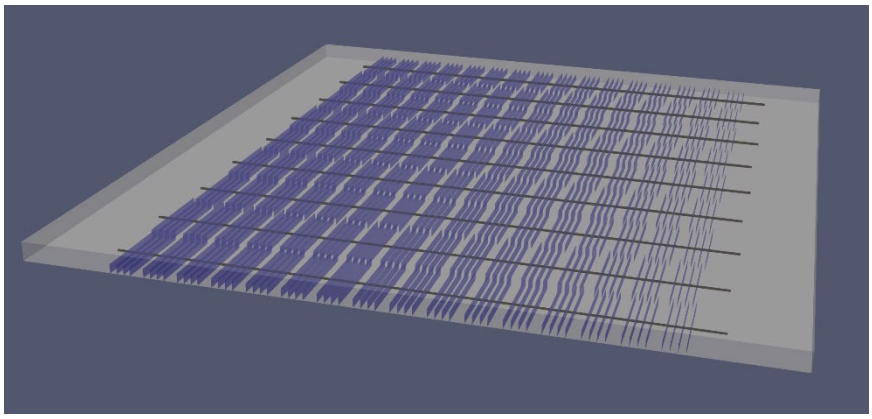
The study is carried out with the following assumptions: (1) The hydraulic fracture geometry is bi-wing. (2) The reservoir is a cuboid. (3) The influence of the fractures from neighboring wells is insignificant at the studied distance. (4) The uncertainties of the hydraulic fractures of each well follow the same distribution as the history matched well.

With the help of EDFM preprocessor, it is easy to add multiple wells, hydraulic fractures, and even complex natural fractures to a reservoir model. Figure 5.6 shows an example of the reservoir model that has multiple wells with hydraulic fractures and

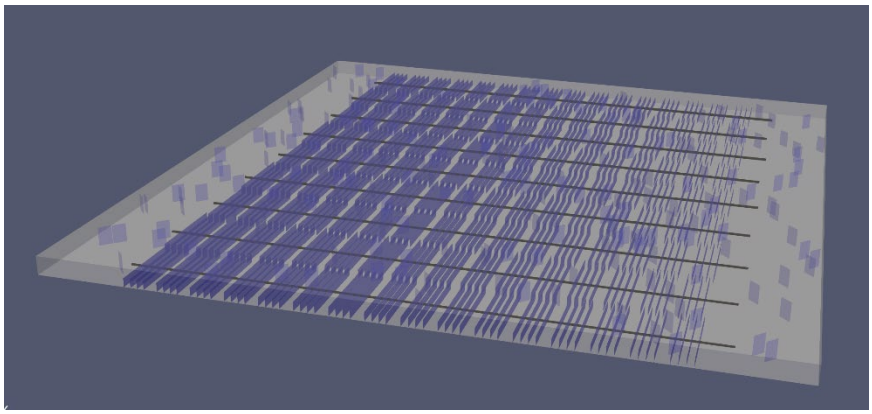
complex natural fractures. EDFM preprocessor can also make it easy to place the well and fractures according to the fracture diagnostic information collected from the field. We can also model the percentage of the effective hydraulic fractures indicated by the distributed temperature sensoring (DTS) data or production log data. An example of randomly generated a percentage of effective fractures for each well is shown in Figure 5.7. Also, for the height of hydraulic fractures, we can adjust the ratio above and below the horizontal wellbore. For this study, the same assumption as history matching is kept that all the fractures are effective. The location of the wells and shape of fractures will follow the pattern used in history matching.



(a) Wells placed evenly in a one-mile reservoir



(b) The clusters of hydraulic fractures added



(c) Natural fractures added

Figure 5.6: An example of the reservoir model including (a) multiple wells with (b) hydraulic fractures and (c) complex natural fractures.

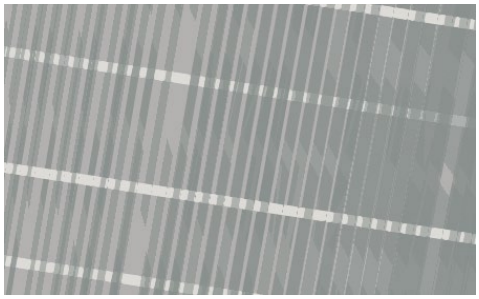
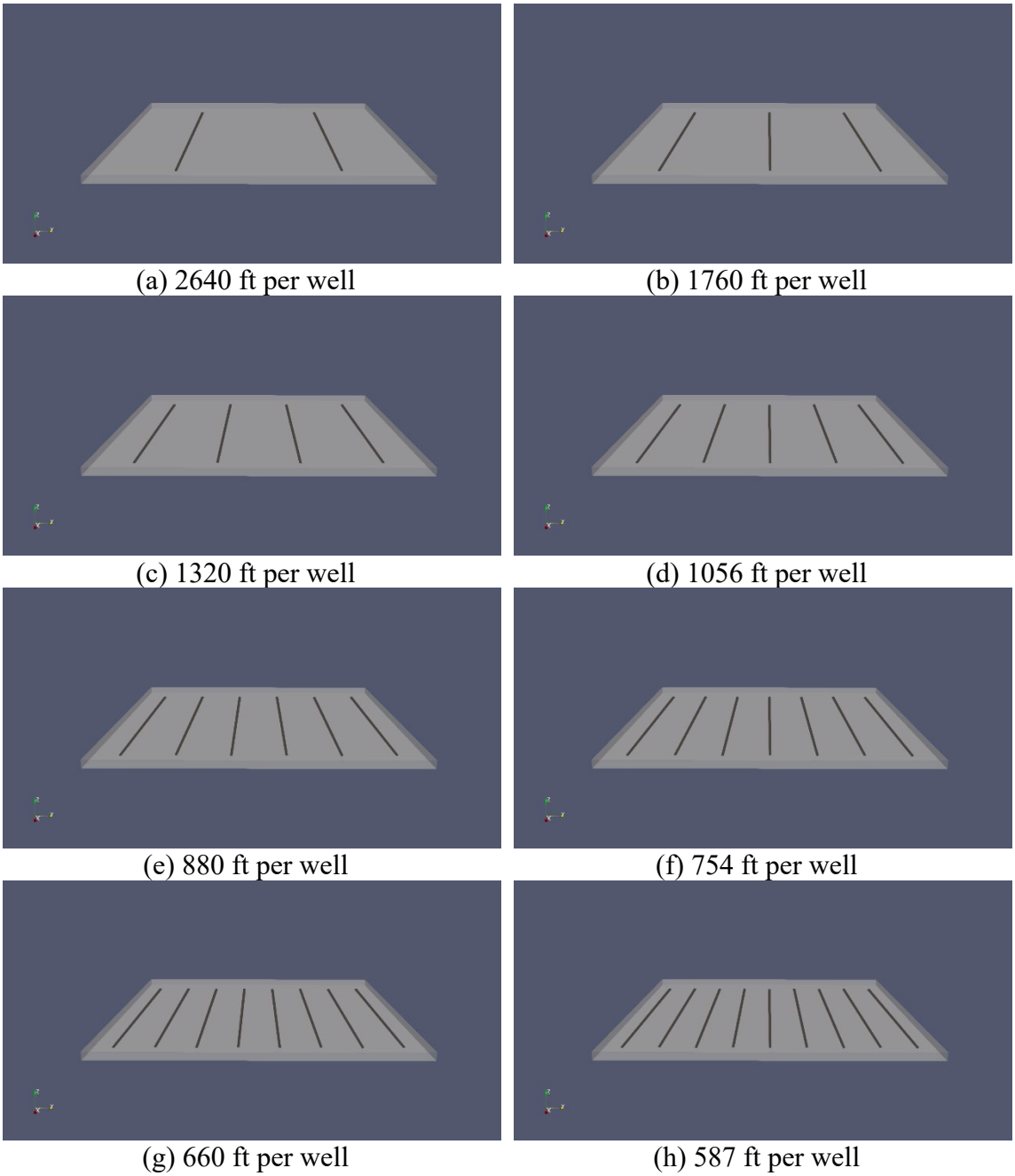


Figure 5.7: An example of the reservoir model with a percentage of effective hydraulic fractures for each well.

For this study, only one layer of horizontal wells and all the wells are used as the producers producing from the same day. An example of the well spacing plan in the reservoir is shown in Figure 5.8. The wells are placed evenly, and the hydraulic fractures are following the zigzag pattern for the maximum drainage efficiency. The center of the horizontal laterals is in the middle of the lower layer, which is identical to that of the history matched well. The length of the wells, the length of the reservoir, and the thickness of the reservoir are also identical to that of the history matching. Different well spacing is realized by placing a different number of horizontal wells on a reservoir with 1-mile width. The detailed uncertain properties of the matrix and fractures are to be discussed in the next section. In this example, we are testing 2-9 horizontal wells in 1 mile wide.



(a) 2640 ft per well (b) 1760 ft per well

(c) 1320 ft per well (d) 1056 ft per well

(e) 880 ft per well (f) 754 ft per well

(g) 660 ft per well (h) 587 ft per well

Figure 5.8: Well spacing plan with different numbers of horizontal wells in the reservoir.

### 5.3.2 Scenario Design

When it comes to the scenarios design for different reservoir realizations, there are four different approaches representing four types of understanding about the problem. The pros and cons of each approach are listed behind.

(1) MCMC is used to generate a large number of the sample that can represent the real distribution of uncertainty. For each well spacing plan, we simulate each of the reservoir realizations and then carry out the analysis. This approach is more statistically reasonable but will take too much computation time. It will only be practical when we can reduce the time for simulation by reliable sector model, or faster machine with more cores.

(2) For each well spacing plan, we can simulate with a wanted number of history matching solution, and compare the calculated NPV directly. It also requires a large number of simulation, and it cannot provide reliable distribution of NPV directly. The benefit of this approach is that simulations we run are more reliable since there is no resampling process and the range of NPV values provides both the most aggressive and conservative prediction.

(3) We generate a wanted number of scenarios from the history matching solution and then run simulations for the combination of those scenarios and well spacing plan. After evaluating all the NPV, we build a proxy model for NPV. We generate another sample with MCMC that can be representative of the real distribution of uncertainties. We apply the NPV proxy model on it for a proxy predicted the real distribution of the NPV for each well. This approach can be regarded as a combination of the first and second approaches. The proxy model is used not only to generate the distribution of uncertainties but also to predict the NPV of each scenario. This method is reliable if when a reliable proxy model is provided.

(4) We cluster the history matched solutions to a wanted several groups and select the best match of each cluster. Then, we test the well plans for each best match. This method is a good choice if we do not want to run a large amount of simulation.

There are also other methods to select the scenarios for simulation, yet we only choose method 2 as a demonstration that has reasonable computational cost. For this case, we have 8 well spacing plans, and for each plan, 50 scenarios of the uncertainty are tested.

The scenarios we ran during history matching are more or less concentrated on the solution regime, and thus we need extra attention when we want to avoid too much bias. All the scenarios designed during the history matching process are presented in Figure 5.9. In this parallel coordinate plot, the color of the line is assigned according to the matching error, RMSE, and the yellow represents the lowest error. If we directly choose the best 50 scenarios, the solutions will be very similar to some extent as shown in Figure 5.10.

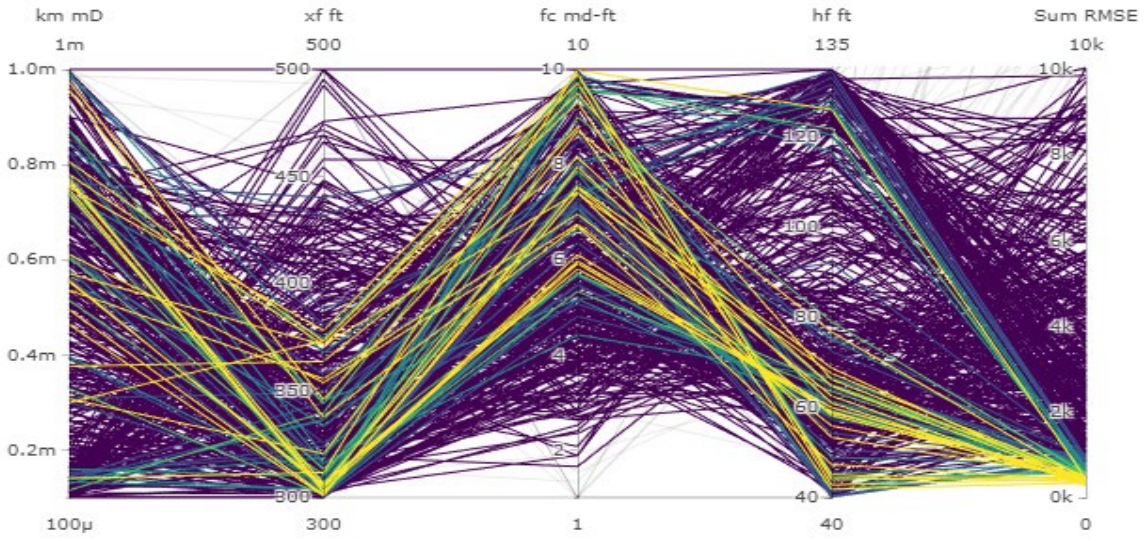



Figure 5.9: All the scenarios evaluated by the simulator during the history matching process.

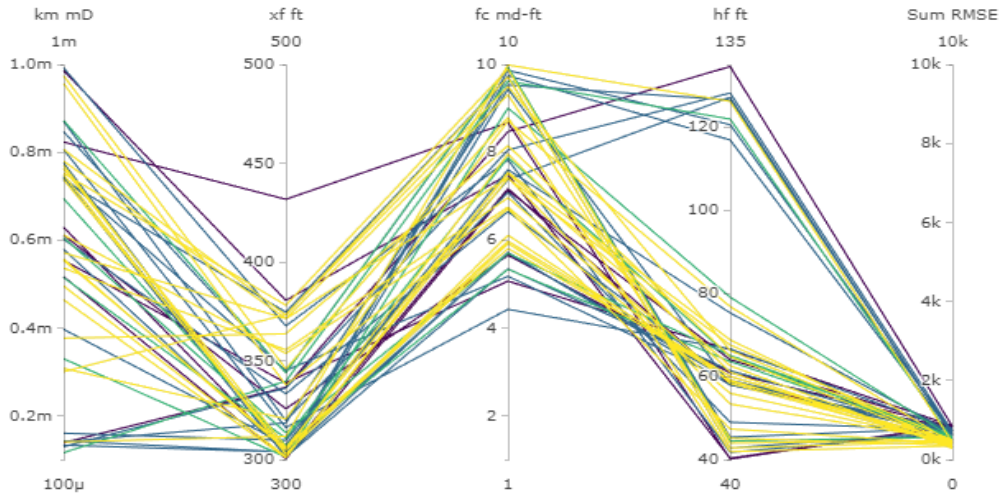
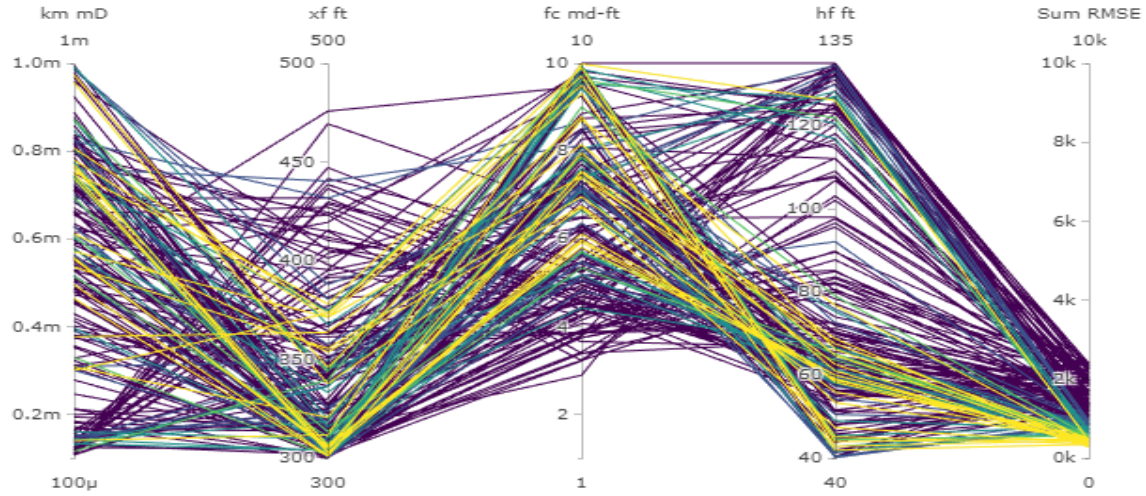
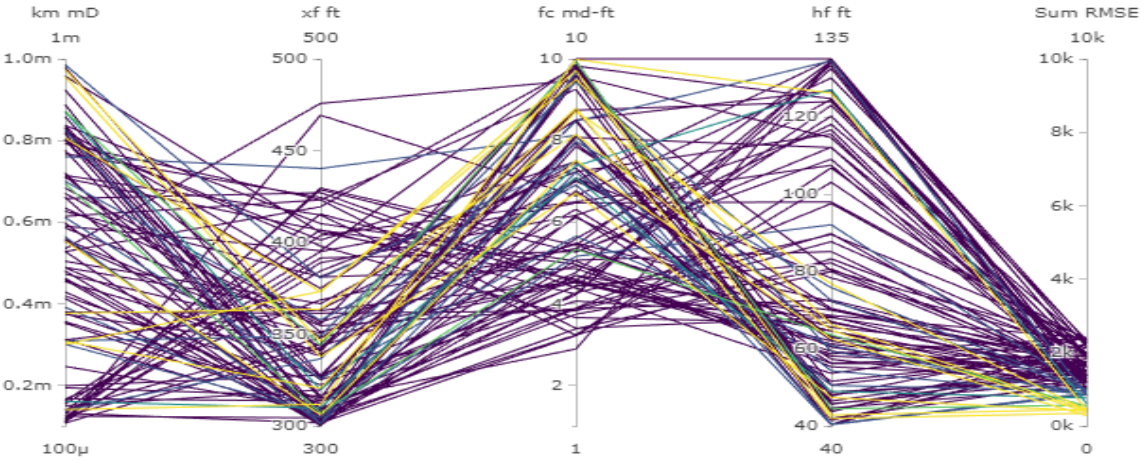



Figure 5.10: Best 50 history matching solutions.

Since diversified solutions are needed to show us more potential reservoir, better coverage of the uncertain space is wanted. Here again, we use the notion of sampling unite to control the diversity of solutions. From all the scenarios, 200 scenarios that have the lowest RMSE are selected. Then, from those 200 scenarios, 120 diversified solutions are selected. Moreover, the 50 with the lowest RMSE are chosen from those 120 scenarios. The number used here is not supported by any theory but a comparison of the scenario designing results. The diversity and overall RMSE is balanced during the comparison. The scenarios selected after each step are shown in Figure 5.11

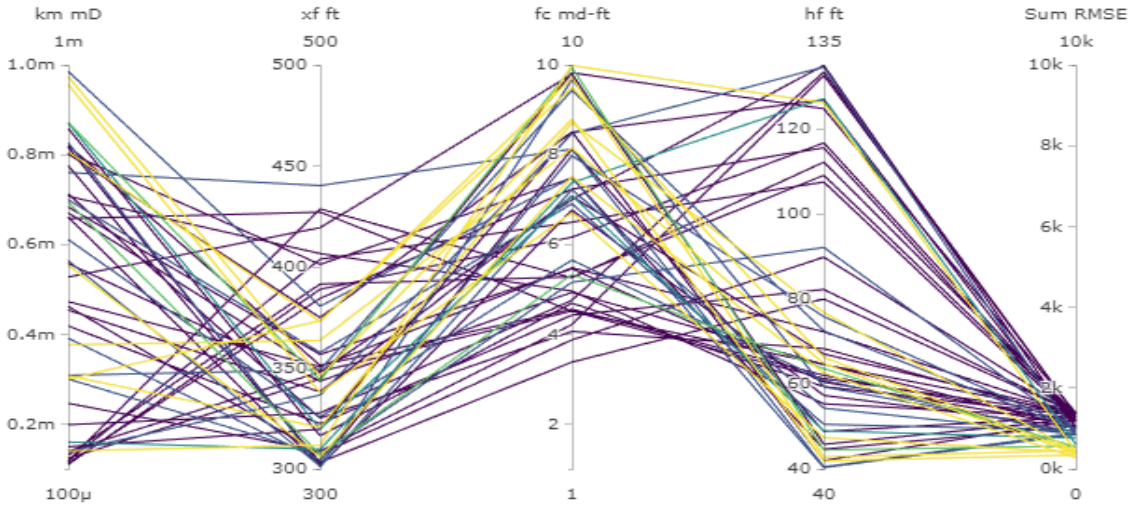



(a) Scenarios selected after the first sampling unit




(b) Scenarios selected after the second sampling unit

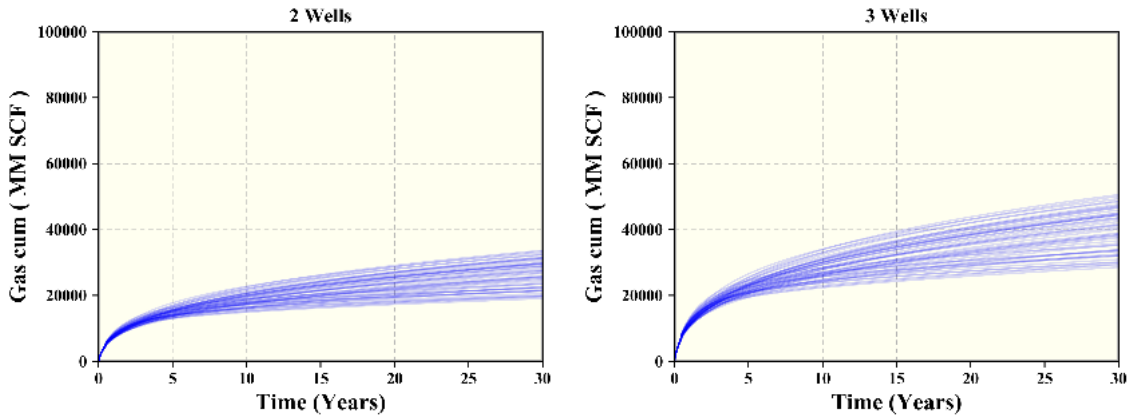
Figure 5.11: Scenarios selected after the (a) first, (b) second and (c) third sampling unit.




(c) Scenarios selected after the last sampling unit

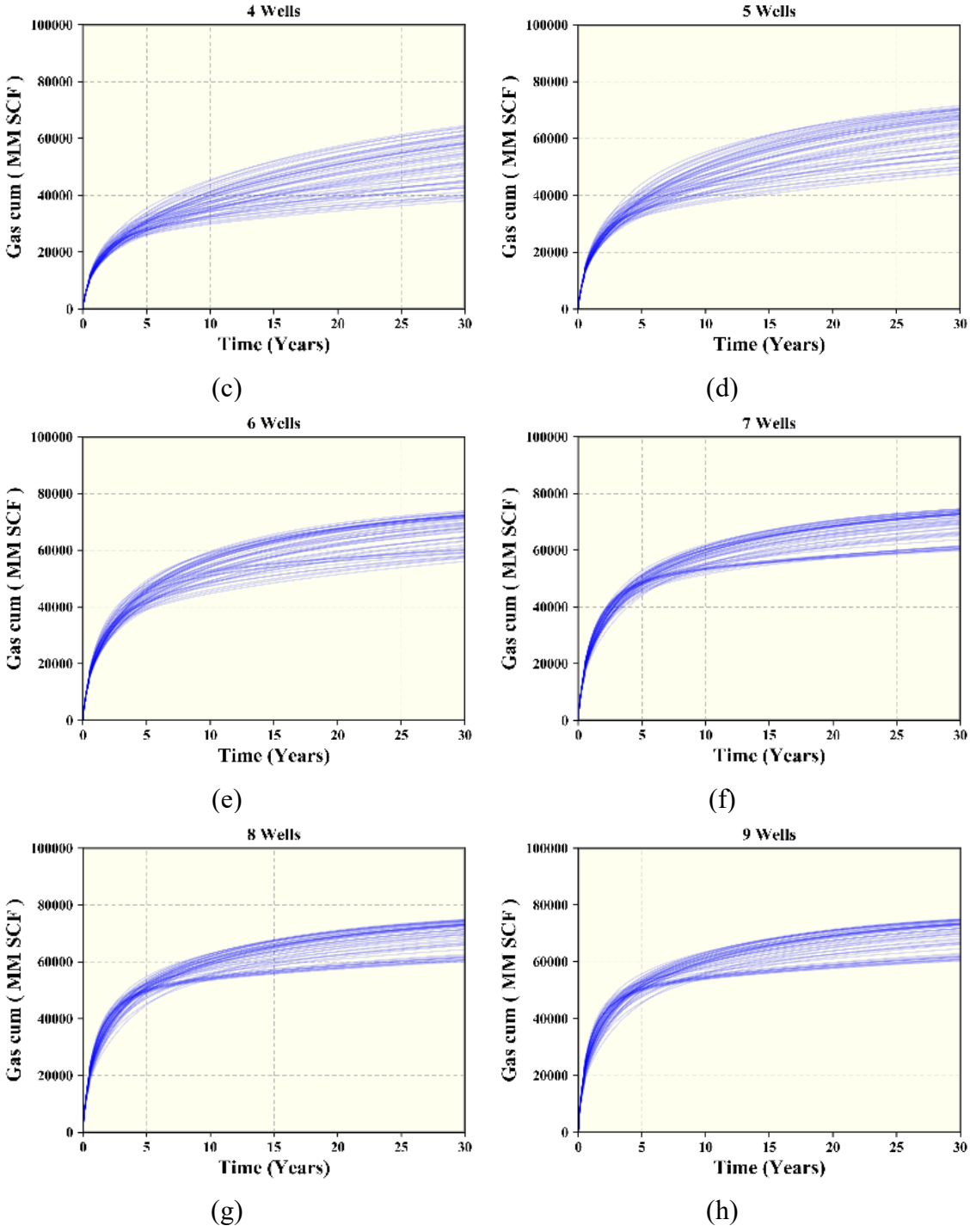
Figure 5.11 continued

Then, the gas production of the field with different wells per mile are plotted in Figure 5.12.




(a) (b)

Figure 5.12: Cumulative gas production of 50 different scenarios when (a) 2, (b) 3, (c) 4, (d) 5, (e) 6, (f) 7, (g) 8, (h) 9 wells are placed and (i) all the combination of scenarios and well spacing plans are plotted.

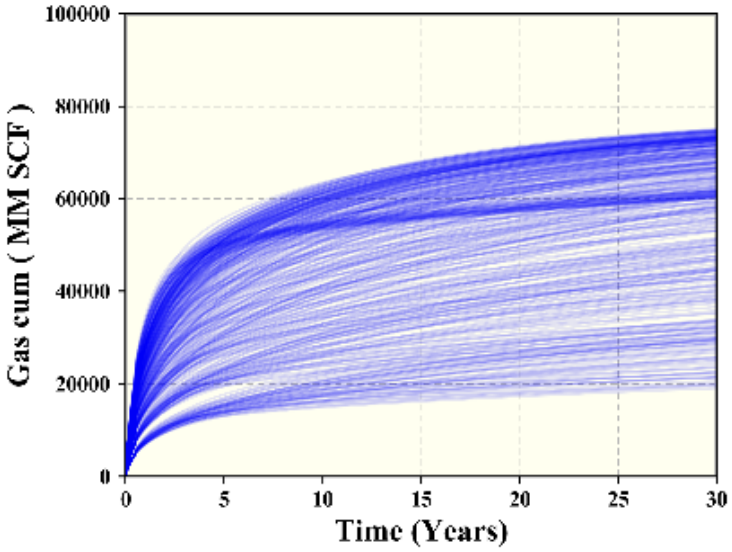



(c) (d)

(e) (f)

(g) (h)

Figure 5.12 continued




(i)

Figure 5.12 continued

### 5.3.3 NPV results

The NPV values are calculated based on the economic environment described in Table5.3.

Table 5.3: Economical environment used in the NPV calculation

| Variable | Value | Unites |
|---|---|---|
| **Well Cost** | 6,000 | M$ |
| **Gas Price** | 3.00 | $/mcf |
| **Water disposal** | 0.55 | $/bbl |
| **Maintainace** | 20,000 | $/Month |
| **Oil Tax** | 5.00 | % |
| **Gas Tax** | 8.00 | % |
| **Other Tax** | 2.00 | % |
| **Annual Discount Rate** | 10 | % |

The NPV of the entire reservoir and the NPV per well of different reservoir realizations are shown together through the box plot in Figures 5.13 and 5.14, respectively.

For this example, we can see that the production per well decrease noticeably when there are more than five wells placed in a mile. Moreover, NPV decreases noticeably when there are more than 6 wells.

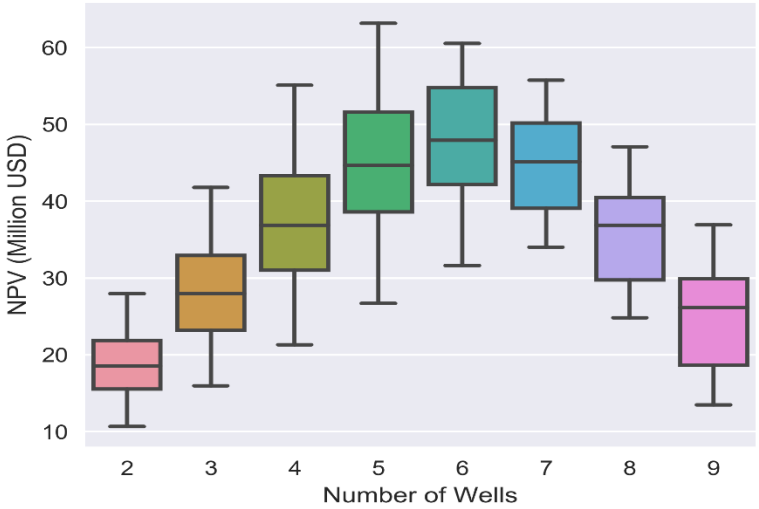



Figure 5.13: Boxplot for total NPV of a 1-mile reservoir when different numbers of wells are placed.

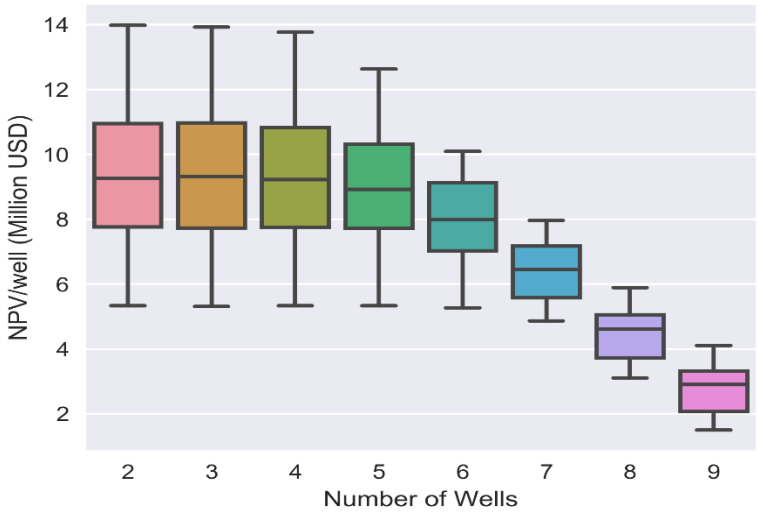



Figure 5.14: Boxplot for NPV per well of a 1-mile reservoir when different numbers of wells are placed.

Also, the plots converted to the well spacings are presented in Figures 5.15 and 5.16, respectively. Based on the boxplot, 880 ft shall be the favorable well spacing if the overall NPV is the target, and 1056 ft shall be favorable if the high NPV per well is the target.

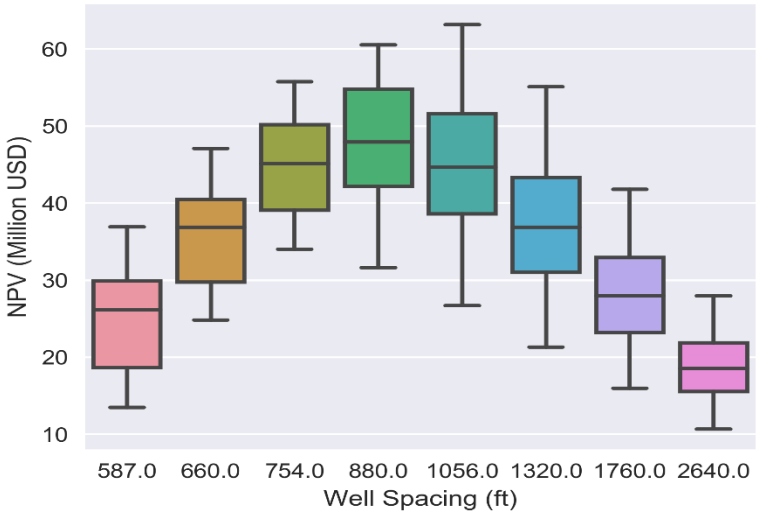



Figure 5.15: Boxplot for total NPV of a 1-mile reservoir when different well spacing is applied.

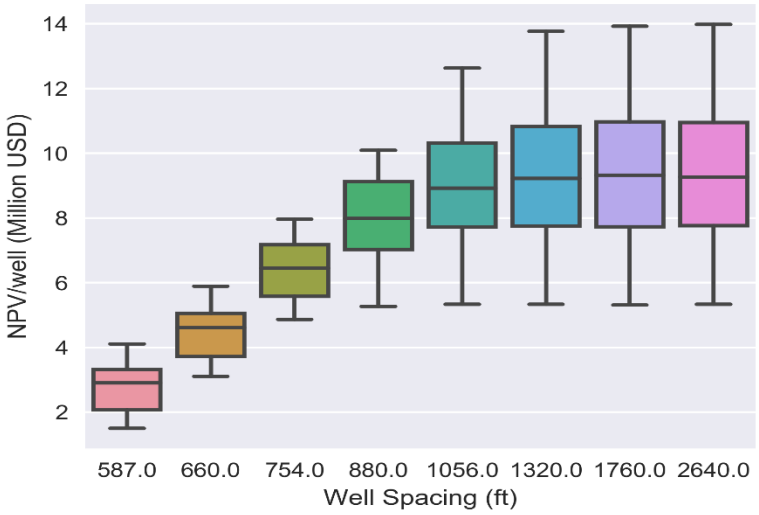



Figure 5.16: Boxplot for NPV per well of a 1-mile reservoir when different well spacing is applied.

Since our workflow did not take the influence of the nearby wells on the extension of hydraulic fractures into account, the real NPV may be lower. Ways to improve the workflow will be discussed in the last chapter of this thesis.

## 5. 4 ADDING NATURAL FRACTURES

The occurrence of natural fractures has the potential to boost the production of shale gas, yet it also makes it difficult to evaluate the efficiency of the completion. If we have the descriptions of the natural fractures, then we can include that in our history matching uncertainties and use that during the simulation of well spacing too.

### 5.4.1 Description of the Natural Fractures

The field case used here does not have the corresponding description of natural fractures. To demonstrate the process and effects of natural fractures, a synthetic description natural fracture is made up here. Table 5.4 describes natural fractures that are constant for all the scenarios. Two sets of planner natural fractures that are perpendicular

to each other are to be generated. The number of natural fractures for each set will be identical. Table 5.5 provides the descriptions of uncertain properties of the natural fractures.

Table 5.4: Descriptions of the constant properties of the natural fractures

| **Fixed-Parameter** | **Unit** | **Value** |
|---|---|---|
| **Number of NF set** | - | 2 |
| **NF height** | ft | 132 |
| **NF Theta** | degree | NF1: 45<br>NF2: 135 |
| **NF Dip Angle** | degree | 90 |
| **NF Width** | ft | 0.1 |

Table 5.5: Uncertain property of the natural fractures

| **Uncertain Parameter** | **Unit** | **Min Value** | **Max Value** |
|---|---|---|---|
| **Total number of NF** | - | 50 | 200 |
| **NF Length** | ft | 50 | 150 |
| **NF Conductivity** | MD-ft | 1 | 10 |

### 5.4.2 Effect of Natural Fractures on History Matching

History matching is re-performed with the newly added uncertain natural fractures variables. The workflow is identical to the previous except that more iterations were run to deal with the increased uncertain dimensions. Among 495 scenarios, 121 cases are regarded as the solutions according to the 3000 criteria. The weighted RMSE of all the scenarios evaluated in the simulation is presented in Figure 5.17. Uncertain parameters of all the designed scenarios and the history matching solutions during the history matching workflow are plotted individually in Figure 5.18. The production profiles of the solutions are plotted in Figure 5.19. The distribution of all the scenarios evaluated by the simulator is plotted in Figure 5.20.

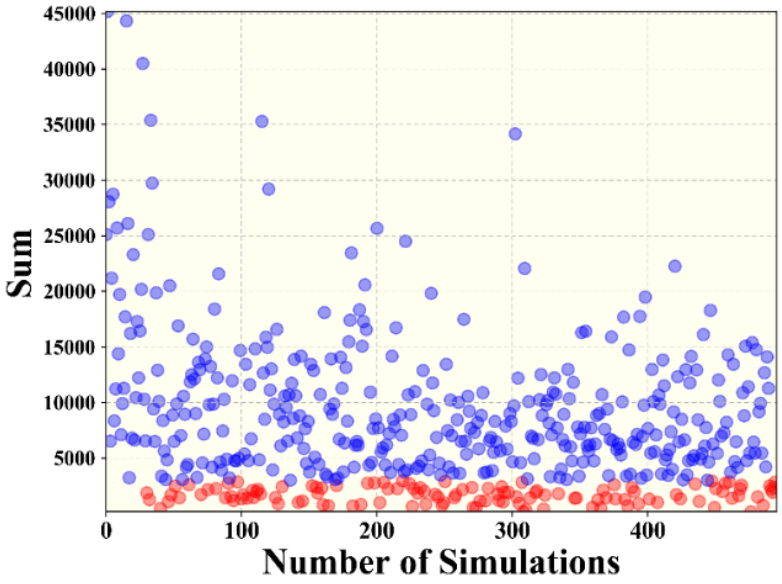
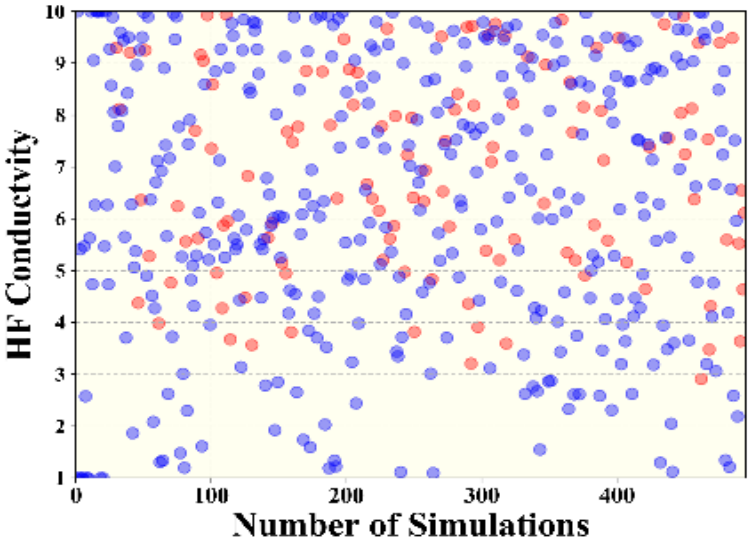
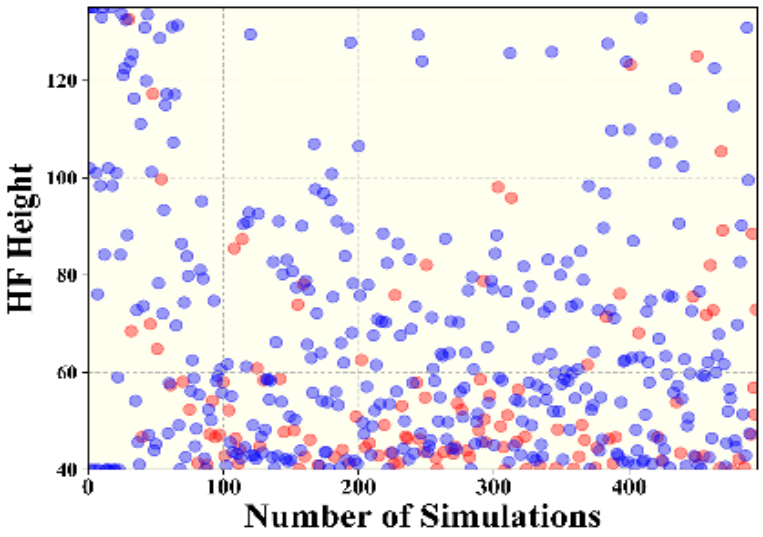



Figure 5.17: Weighted RMSE of all the simulation runs with natural fractures (red dots represent history-matching solutions, and blue dots are non history-matching solutions).

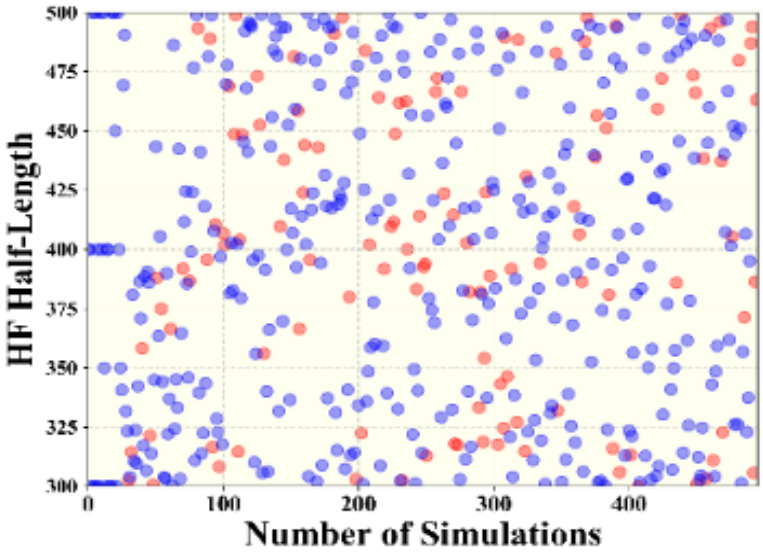



(a) Conductivity of the hydraulic fracture

Figure 5.18: Uncertain parameters of the history matching solutions with natural fractures (red dots represent history-matching solutions, and blue dots are not history-matching solutions).

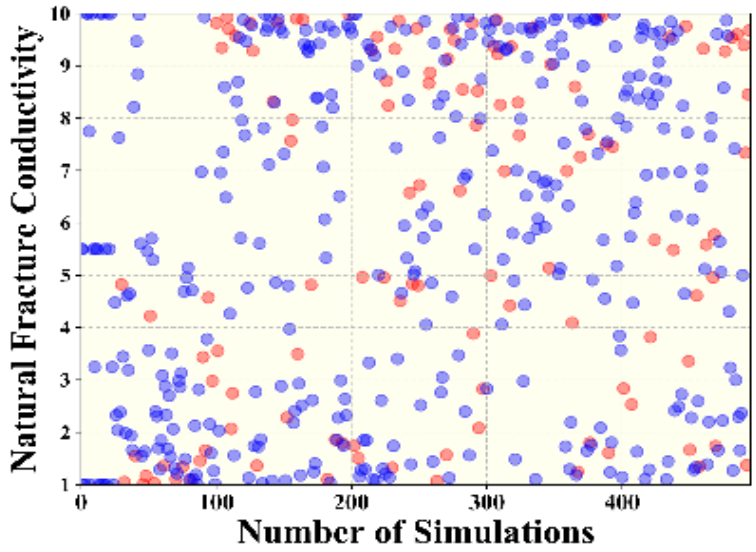



(b) Height of the hydraulic fracture

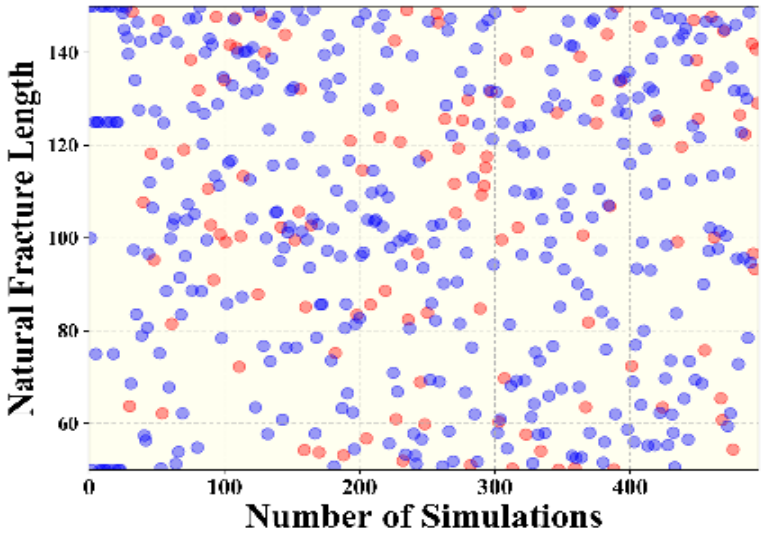



(c) Half-length of the hydraulic fracture

Figure 5.18 continued

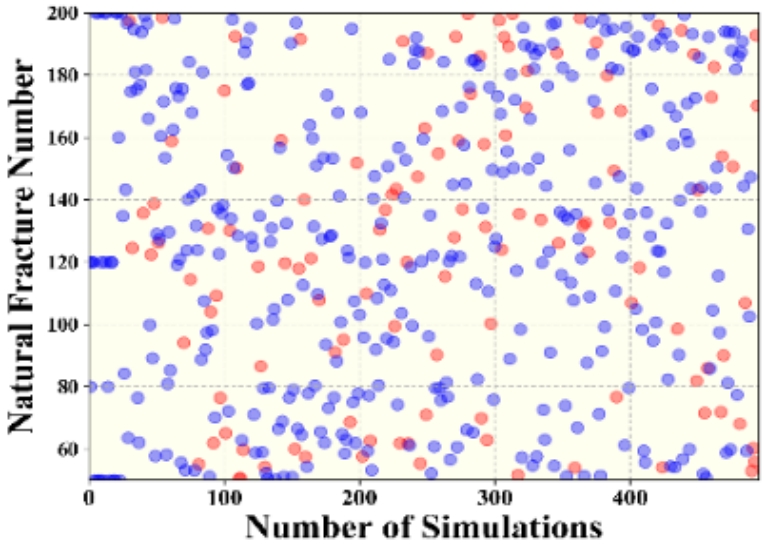



(d) Conductivity of the natural fracture

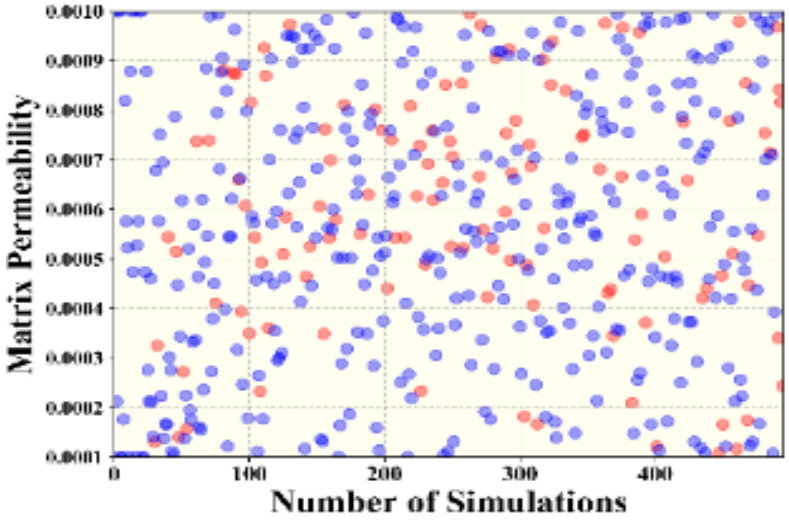



(e) Length of the natural fracture

Figure 5.18 continued

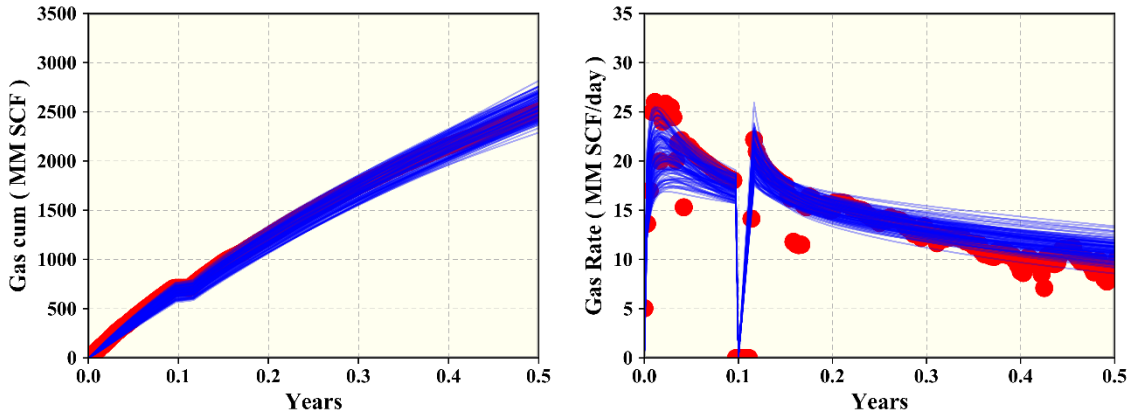



(e) Number of the natural fracture




(f) Permeability of reservoir matrix

Figure 5.18 continued




(a) Cumulative gas production (b) Gas flow rate

Figure 5.19: Simulation results of the history matching solutions from iterative response surface methodology when natural fractures are considered.

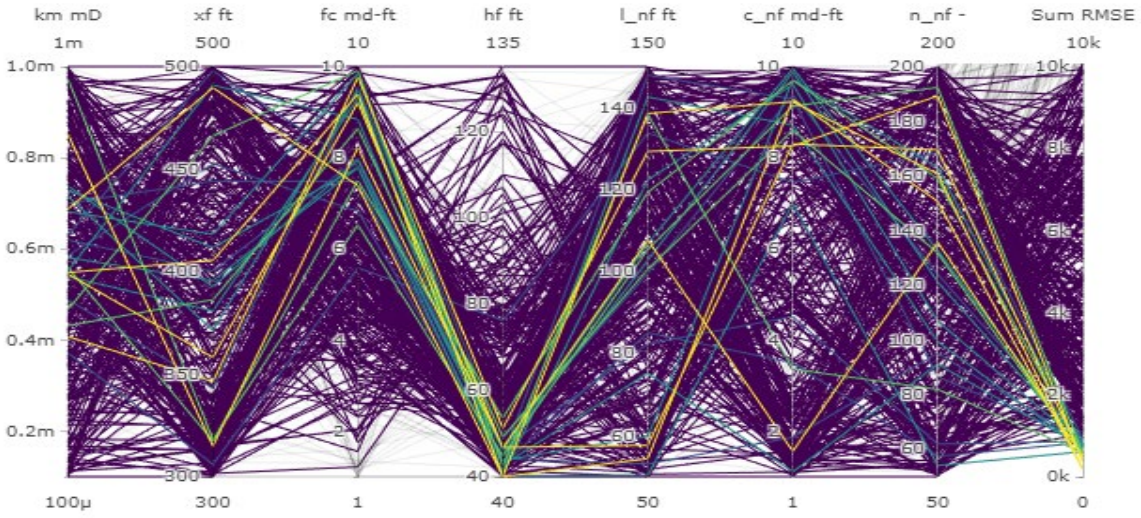
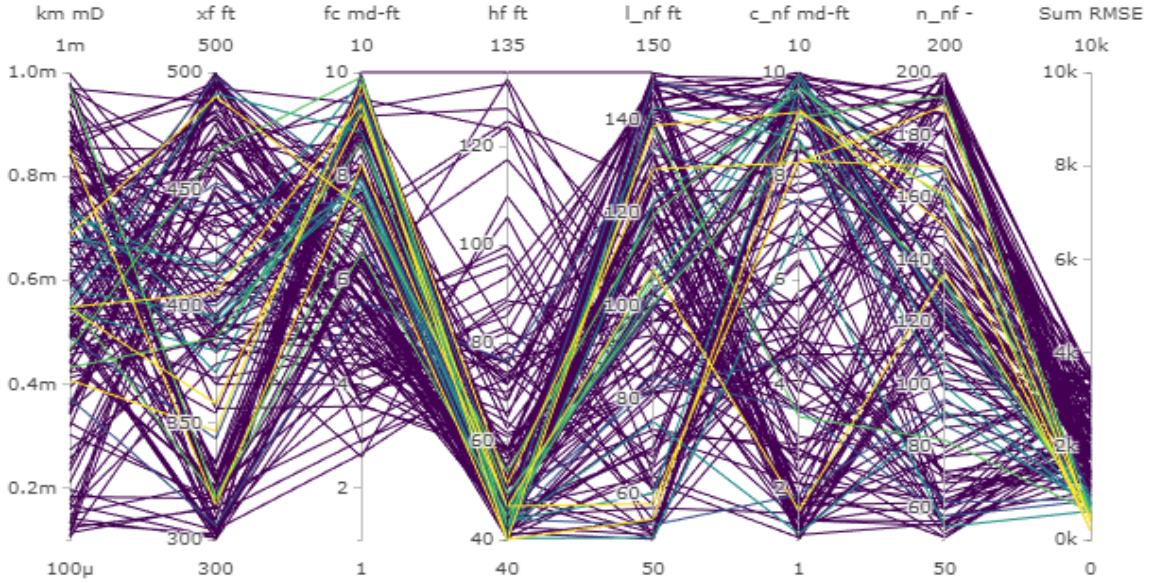



Figure 5.20: All the scenarios evaluated by the simulator during the history matching when natural fractures are considered.

For the well spacing plan, the properties of the natural fractures will be modified according to the size of the reservoir. Here we only modify the number of natural fractures to keep the same fracture density that is calculated by the equation below.

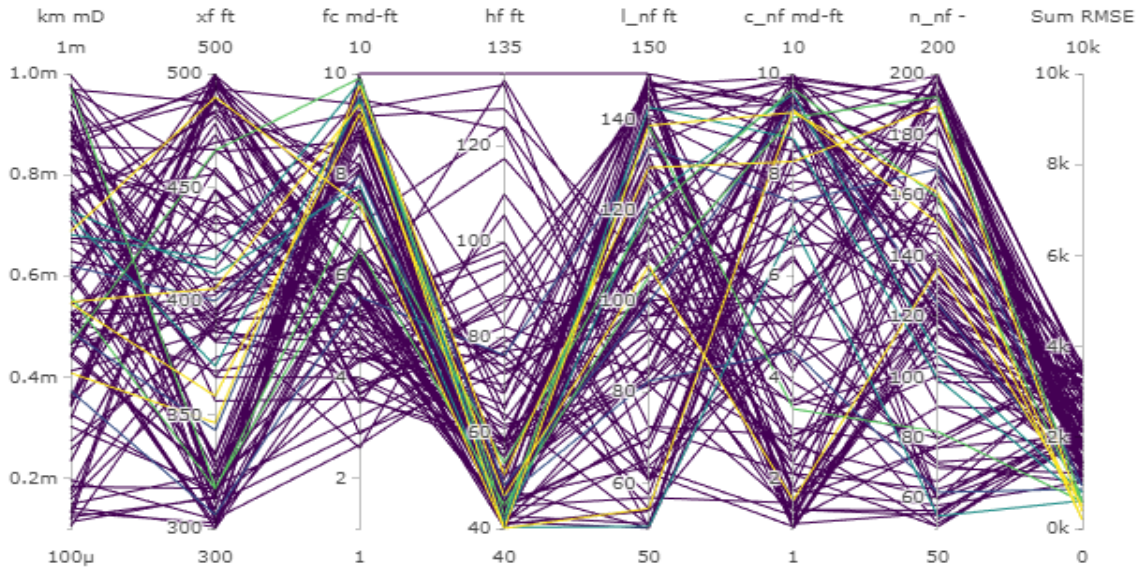
$$fracture\ density = \frac{surface\ area\ of\ fractures}{volume\ of\ fractured\ matrix}. \quad (5.1)$$

The process to select the scenarios for well spacing plans is shown in Figure 5.21, and the gas production of the different well spacing plan with different reservoir realizations is shown in Figure 5.22.

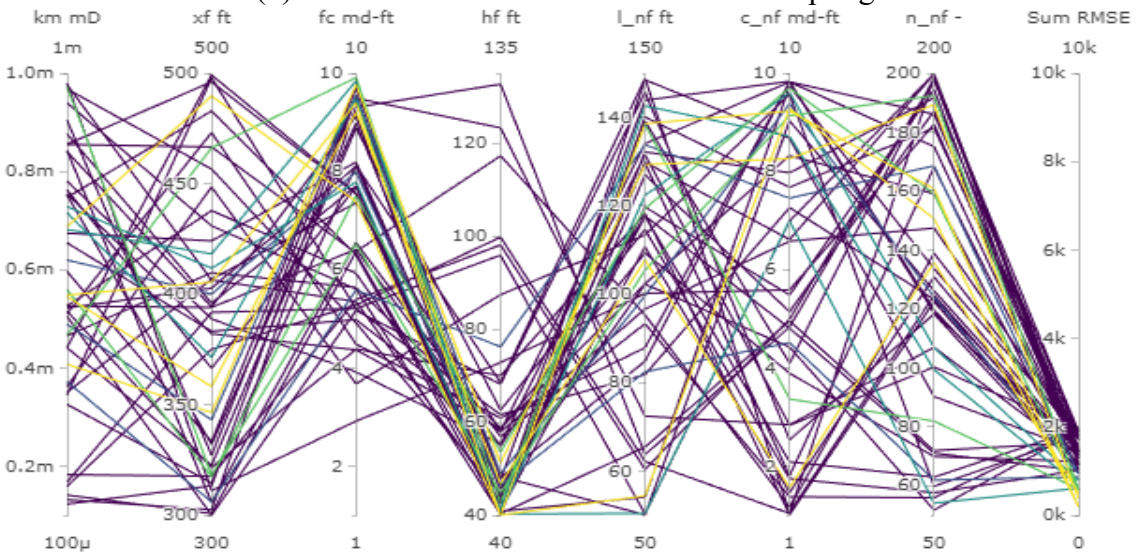



(a) Scenarios selected after first sampling unit

Figure 5.21: Scenarios selected after the (a) first, (b) second and (c) third sampling unit when natural fractures are considered.




(b) Scenarios selected after the second sampling unit




(c) Scenarios selected after the third sampling unit

Figure 5.21 continued

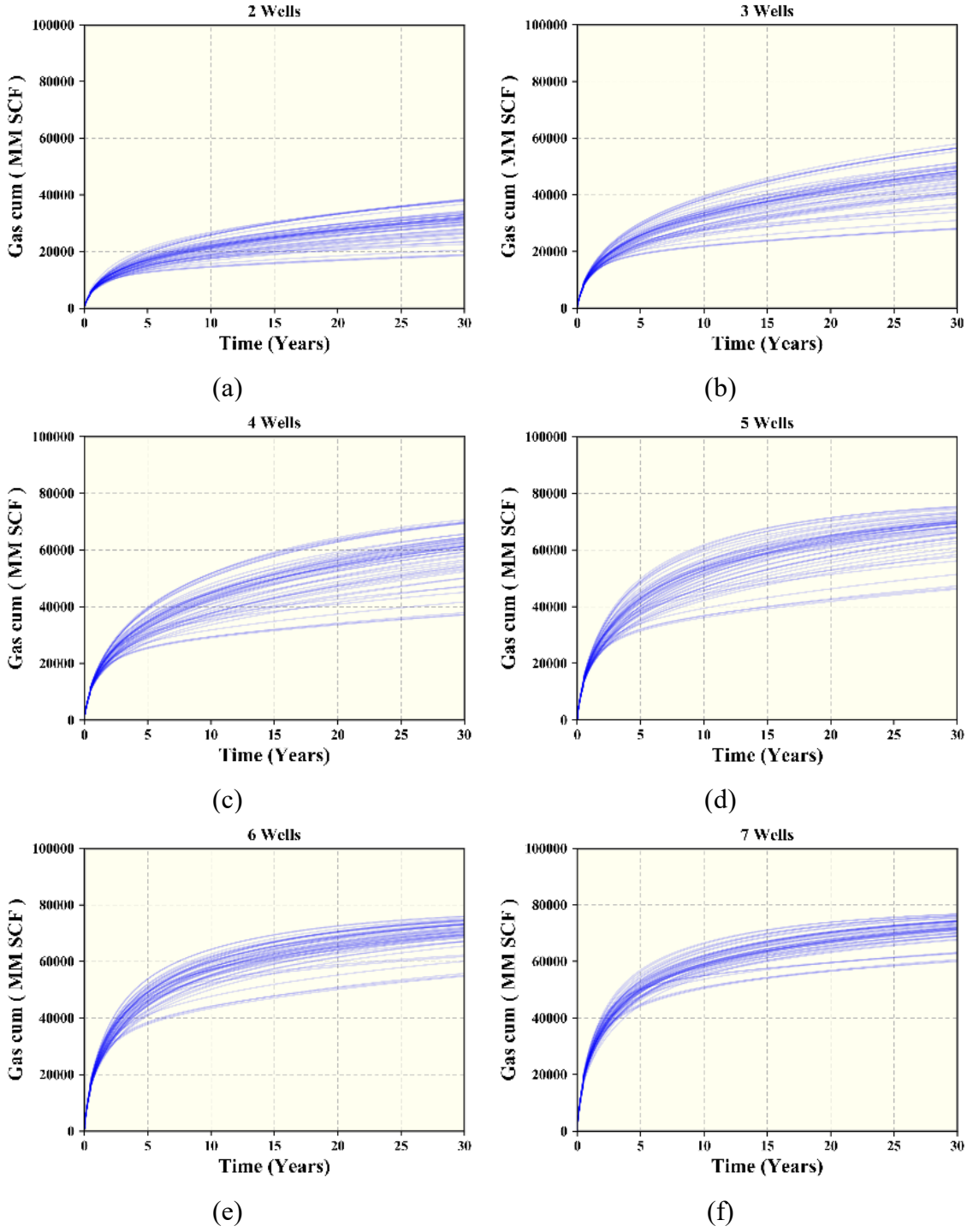
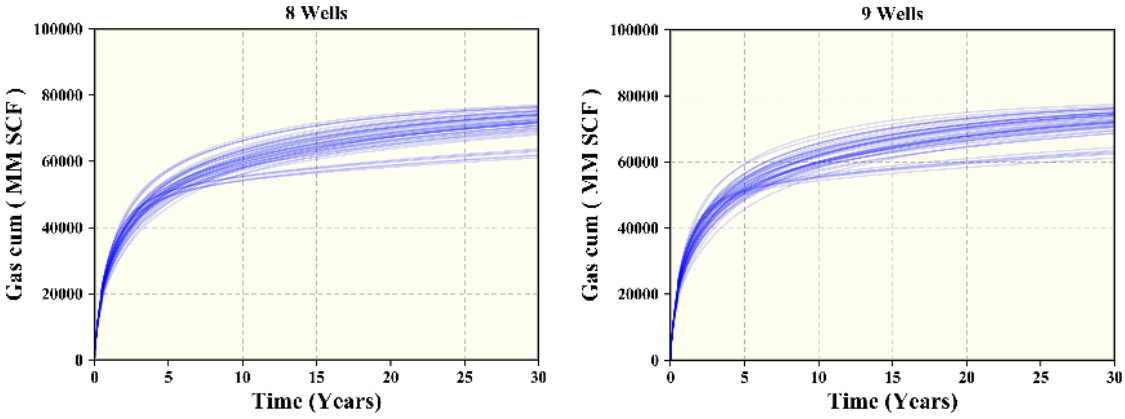
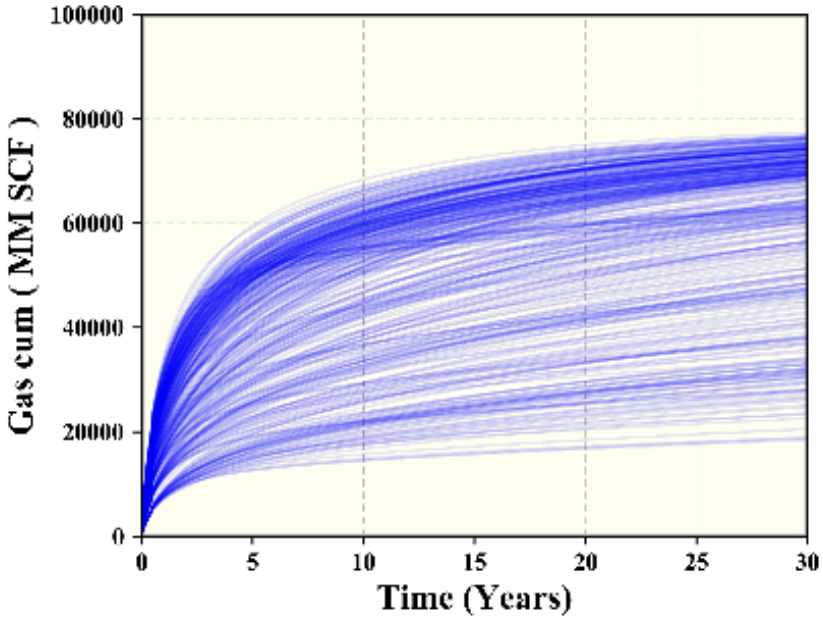



Figure 5.22: Cumulative gas production of 50 different scenarios when (a) 2, (b) 3, (c) 4, (d) 5, (e) 6, (f) 7, (g) 8, (h) 9 wells are placed in naturally-fractured reservoir, and (i) all the combination of scenarios and well spacing plans are plotted.




(g) (h)




(i)

Figure 5.22 continued

### 5.4.3 Comparison Results

The NPV for well spacing plans is calculated with the same economic environments. The corresponding results are shown similarly in Figures 5.23, 5.24,5.25, and 5.26.

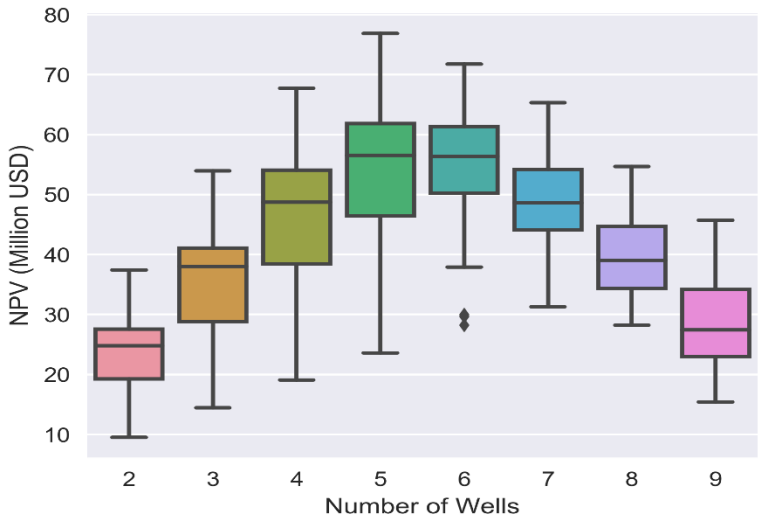



Figure 5.23: Boxplot for total NPV of a 1-mile reservoir when different numbers of wells are placed.

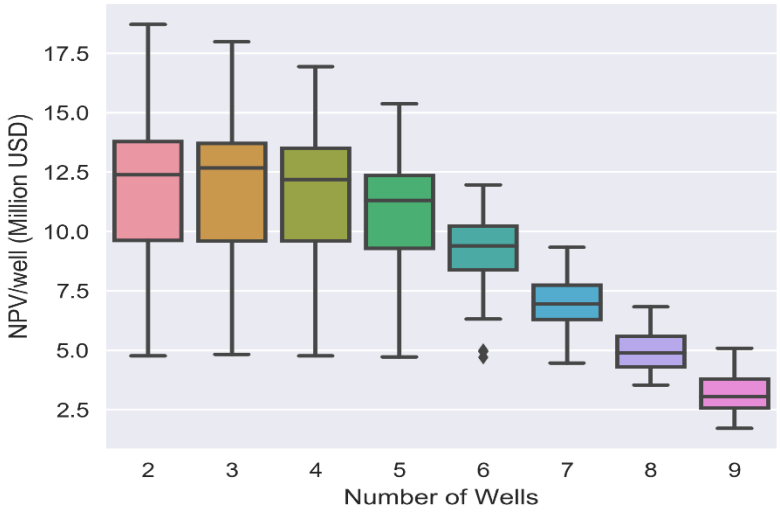



Figure 5.24: Boxplot for NPV per well of a 1-mile reservoir when different numbers of wells are placed.

The overall NPV is slightly higher when natural fractures are included in the workflow. We also notice that the range of potential NPV is wider when the natural fracture

is considered. The turning point for overall NPV and NPV per well is 5 and five which are similar to the previous conclusion and so will be for the well spacing. It predicts more NPV and will affect if we set a critical NPV for the spacing plan.

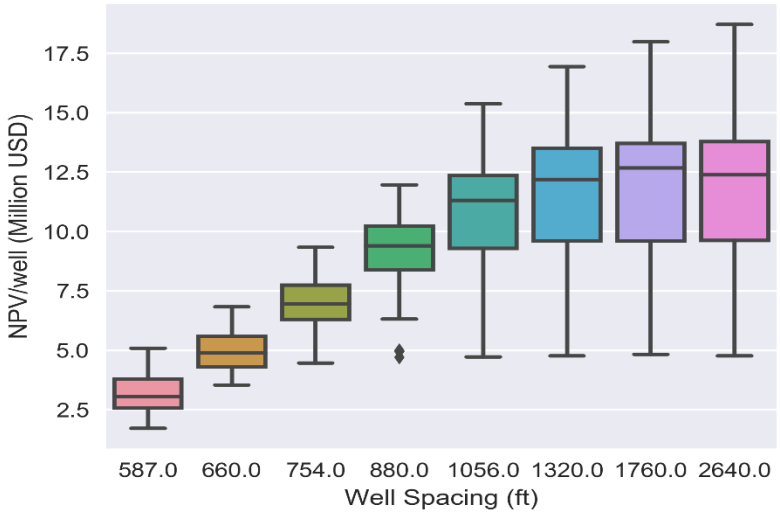



Figure 5.25: Boxplot for total NPV of a 1-mile reservoir when different well spacing is applied.

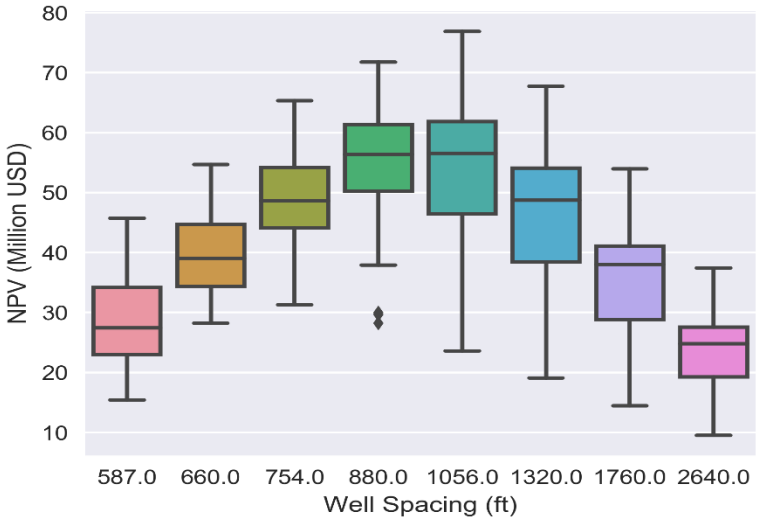



Figure 5.26: Boxplot for NPV per well of a 1-mile reservoir when different well spacing is applied.

## 5.5 CONCLUSIONS

In this chapter, assisted history matching workflow and well spacing workflow are performed on a shale gas reservoir. The platform makes the implementation easier by using a very similar file structure and workflow structure. Based on the assumption of the economic environment and the natural fractures, the key findings are summarized as follows:

1. The method to control the diversity of the solution in history matching is a success. The production history is matched, and the uncertain space is covered considerably well. Although the ratio of history matching solution to all the designed scenarios is not very high, the information gathered can be enough for further analysis.
2. The uncertainty quantification is not performed following the traditional scheme. The marginal distribution of individual uncertain properties is not calculated with MCMC or another method. It is because they are not necessary for the well spacing workflow chosen. An example that quantified the uncertainty is introduced in the next chapter.
3. The well spacing workflow in this thesis does not take the geomechanics into account, and thus the uncertainties gathered through history matching workflow is directly used. The simulation of the different well spacing plan with different reservoir realization is applied to a 1-mile-wide reservoir to compare the total NPV and NPV per well. The suggestion of well spacing is given according to the targets of the project.
4. The occurrence of natural fractures will affect the history matching results and thus the long-term production. In this case, natural fractures promote the NPV

of the assets. Assessing the density of natural fractures is helpful when we want to evaluate the potential value.

# Chapter 6: Automatic History Matching with a Gas Condensate Reservoir

In this chapter, we are utilizing the platform to perform automatic history matching with a gas condensate well. The basic reservoir model is prepared based on the field and lab reports, while some of the hard-to-measure properties of the model are regarded as uncertain. To reduce the number of resources necessary for the history matching process, we performed a few studies including sensitivity analysis and model simplifications. After presenting the conventional history matching results, several approaches to dig more value from the simulations are provided. MCMC is used as a more reliable way to estimate the posterior distribution of uncertainties and productions. Clustering is used to reduce the number of resources necessary for the further simulation study.

## 6.1 Reservoir Model

The reservoir model is build based on the descriptions provided by the reports from the operator, and some of the properties are regarded as uncertain. To deal with the possible phase behavior of the gas condensate reservoir, the fluid properties are matched to the lab experiments for the compositional model.

### 6.1.1 Reservoir Description and Uncertainties

The case study is a shale gas condensate well in Duvernay Shale Formation, Canada that has initial reservoir pressure of 8601 psi. A horizontal well was drilled and completed with 27-stage fracture stimulation. In each stage, four perforation clusters were shot. The production profiles of bottom-hole pressure (BHP), condensate rate, gas rate, and water rate are recorded for 613 days. The goal of this study is to history match these production data and provide a forecast of future production with reservoir simulation. The first part of the simulation is constrained with the actual gas production rate, and for the production forecast period, BHP is constrained at 500 psi until the 30 years. Table 6.1 summarizes the assumption of constant input parameters for the reservoir model.

Table 6.1: Reservoir parameters in the simulation model for the horizontal well

| Reservoir Description | Value | Unit |
|---|---|---|
| **Model dimension (x × y × z)** | 6130 × 2000 × 132 | ft |
| **Number of grid blocks (x × y × z)** | 613 × 201 × 13 | - |
| **Grid cell dimension (x × y × z)** | 10 × 10 × 10.15 | ft |
| **Initial reservoir pressure** | 8601 | psi |
| **Reservoir temperature** | 239 | $^{o}$F |
| **Horizontal well length** | 6127 | ft |

Uncertainties of the reservoir and fractures can be unlimited, but we will focus on the parts that are making the difference. Several assumptions and simplifications are made. (1) The reservoir is assumed to have homogeneous permeability and porosity for simplification. (2) The Langmuir isotherm is used to describe the gas desorption in the organic matter of the formations. (3) The well is assumed to be in the middle of the reservoir with equal distance to the top and bottom. (4) The fracture pattern is assumed to be multiple bi-wing fractures with evenly fracture spacing, but the number of effective fractures is uncertain**.** The pattern of the model is shown in Figure 6.1.

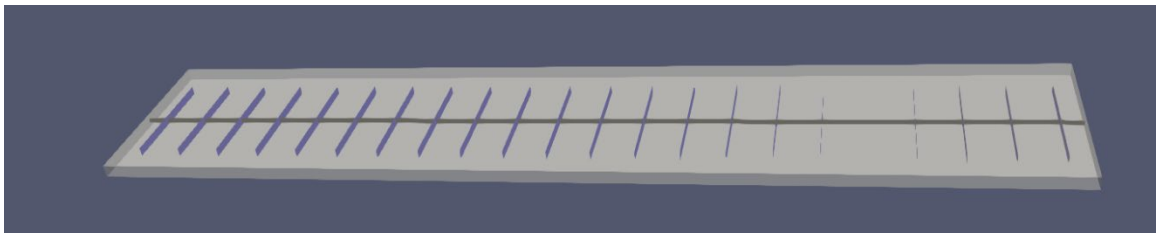


Figure 6.1: Simulation model for the horizontal well with multiple fractures. The black line in the middle is the horizontal well; the blue surfaces vertical to the well are hydraulic fractures.

In this study, the range of 12 uncertain parameters is listed in Table 6.2. They are assigned based on prior knowledge about this well. Uniform distributions are assumed for all of these uncertain parameters since only the minimum and maximum, but not the most likely values, are known.

Table 6.2: Summary of the ten uncertain parameters and their prior distributions

| Code | Uncertain parameter | Unit | Min | Max |
|---|---|---|---|---|
| **A** | Matrix permeability | mD | 0.00002 | 0.0004 |
| **B** | Hydraulic fracture half-length | ft | 100 | 700 |
| **C** | Hydraulic fracture height | ft | 20 | 100 |
| **D** | Hydraulic fracture conductivity | md-ft | 1 | 20 |
| **E** | Equivalent fracture length | - | 0.01 | 0.3 |
| **F** | No. of effective hydraulic | - | 13 | 27 |
| **G** | Matrix porosity | - | 0.02 | 0.15 |
| **H** | Fracture water saturation | - | 0.05 | 0.95 |
| **J** | Langmuir inverse pressure | 1/psi | 0 | 0.002 |
| **K** | Rock compressibility | 1/psi | 0.000001 | 0.00001 |
| **L** | Matrix water saturation | - | 0.1 | 0.8 |
| **M** | Langmuir volume | gmol/lb | 0.023 | 0.091 |

Table 6.3: Summary of the response parameters and their weights to calculate RMSE

| | Unit | Weight |
|---|---|---|
| **Oil Rate** | MSTB/day | 5 |
| **Gas Rate** | MMSCF/day | 20 |
| **Water Rate** | MSTB/day | 5 |
| **Cumulative Oil Production** | MSTB | 5 |
| **Cumulative Gas Production** | MMSCF | 20 |
| **Cumulative Water Production** | MSTB | 10 |
| **BHP** | psi | 30 |

### 6.1.2 Phase Behavior

The compositional simulation is used to characterize the phase behavior of this gas condensate reservoir. The fluid modeling used the Peng-Robinson equation of state, and the regression is made to match the critical properties of the heavy component. The gas condensate is carefully divided into six different pseudo-components, i.e., $CO_2$, $CH_4$, $C_2$, $C_3$, $C_4$-$C_6$, $C_{7+}$, and their corresponding mole fractions are 1%, 72%, 7%, 5%, 7%, and 8%, respectively. The key fluid properties are calculated based on these components using CMG-WinProp (CMG-WinProp, 2016): Fluid density is 20.7947537 lb/ft$^3$, fluid viscosity is 0.0399 cp, saturation pressure 4091.49 psi, critical pressure is 4313.2 psi, and critical temperature is 100.99 $^o$F. The other input data required for the Peng-Robinson equation-

of-state (EOS) are listed in Table 6.3. The binary coefficient used for flash calculation is listed in Table 6.4.

Table 6.4: Compositional data for the Peng-Robinson EOS

| Component | Molar fraction | Critical pressure (atm) | Critical temperature (K) | Critical volume (L/mol) | Molar weight (g/gmol) | Acentric factor | Parachor coefficient |
|---|---|---|---|---|---|---|---|
| $CO_2$ | 0.01 | 72.8 | 304.2 | 0.094 | 44.010 | 0.225 | 78 |
| $CH_4$ | 0.72 | 45.4 | 190.6 | 0.099 | 16.043 | 0.008 | 77 |
| $C_2$ | 0.05 | 48.2 | 305.4 | 0.148 | 30.070 | 0.098 | 108 |
| $C_3$ | 0.07 | 41.9 | 369.8 | 0.203 | 44.097 | 0.152 | 150.3 |
| $C_4$-$C_6$ | 0.07 | 37.9 | 413.41 | 0.290 | 68.278 | 0.221 | 212.6 |
| $C_{7+}$ | 0.08 | 29.2 | 659.09 | 0.556 | 141.190 | 0.377 | 399.5 |

Table 6.5: Binary interaction parameters for oil components

| Component | $CO_2$ | $CH_4$ | $C_2$ | $C_3$ | $C_4$-$C_6$ | $C_{7+}$ |
|---|---|---|---|---|---|---|
| $CO_2$ | 0 | 0.105 | 0.13 | 0.125 | 0.1159 | 0.00718 |
| $CH_4$ | 0.105 | 0 | 0 | 0 | 0 | 0 |
| $C_2$ | 0.13 | 0 | 0 | 0 | 0 | 0 |
| $C_3$ | 0.125 | 0 | 0 | 0 | 0 | 0 |
| $C_4$-$C_6$ | 0.1159 | 0 | 0 | 0 | 0 | 0 |
| $C_{7+}$ | 0.00718 | 0 | 0 | 0 | 0 | 0 |

The pressure-temperature (P-T) phase diagram for the fluid composition was obtained as displayed in Figure 6.1. Another comparison of experiment results and simulation results after lumping are presented in Fig. 6.2.

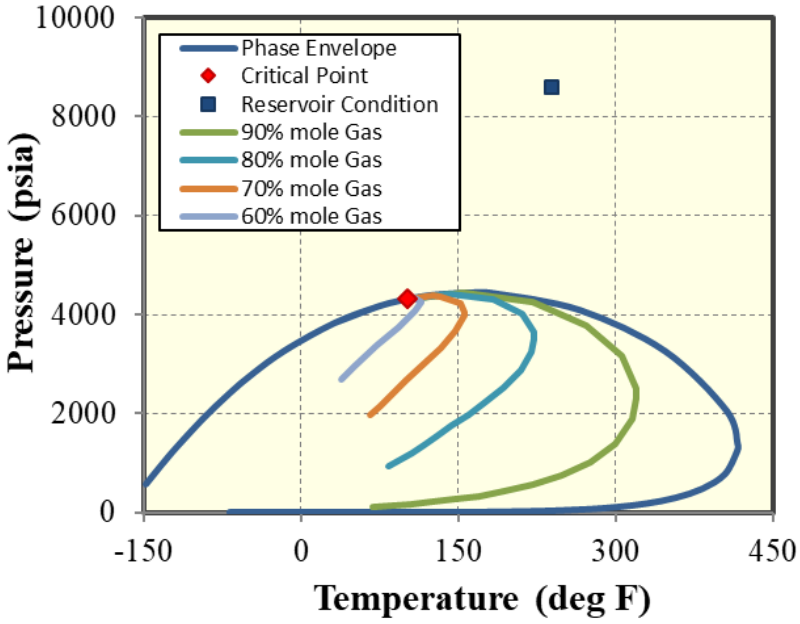
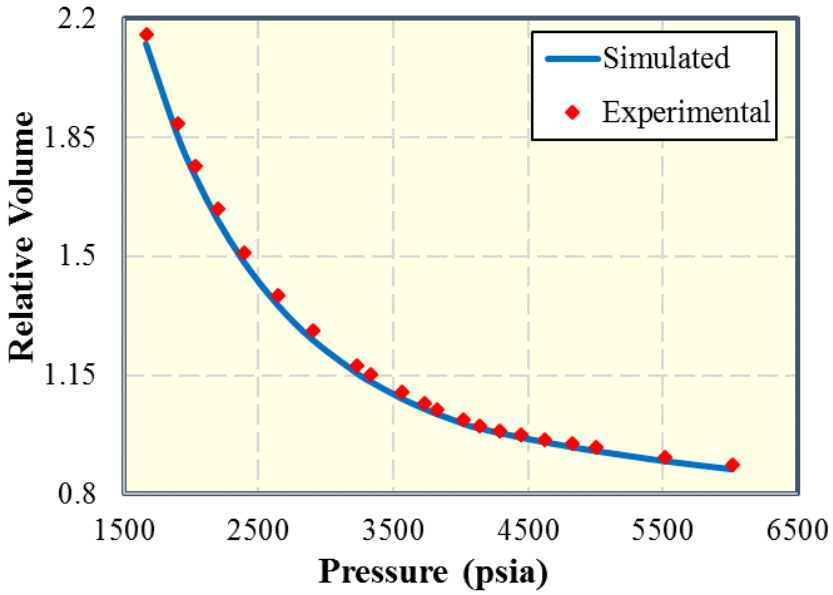
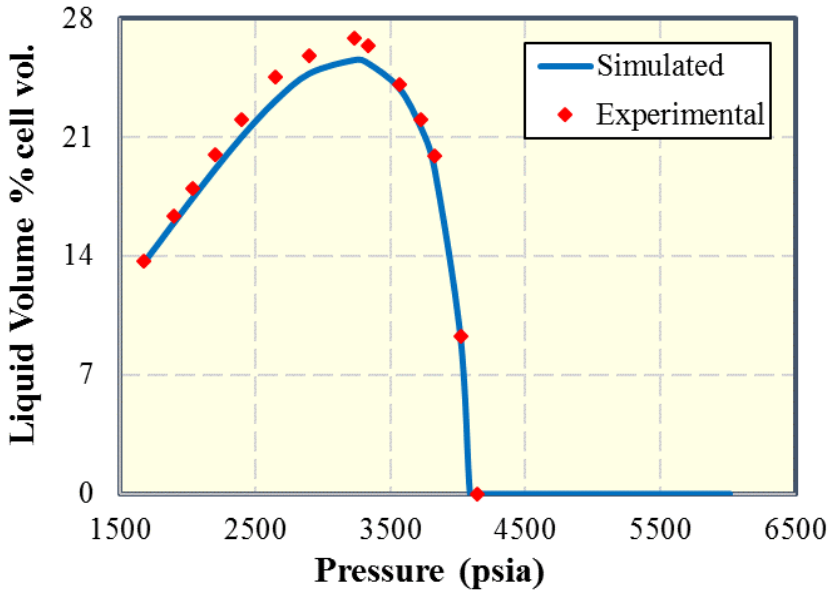



Figure 6.2: Two-phase envelop of the reservoir fluids.

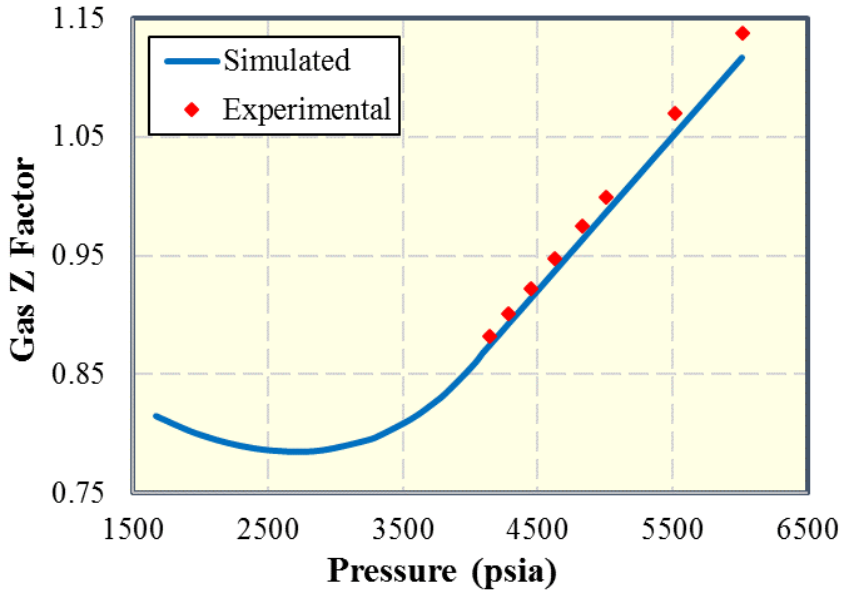



(a) Relative volume

Figure 6.3: Matched fluid model against the lab measurement results.

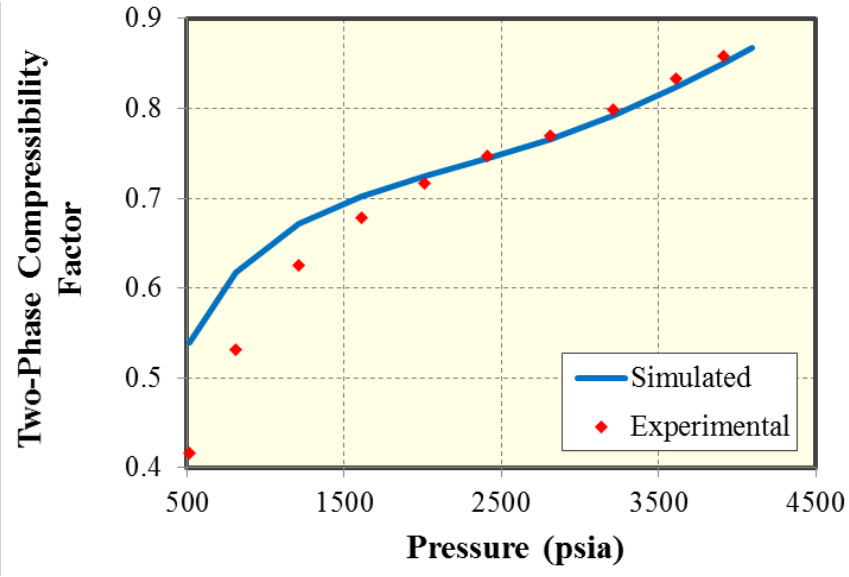



(b) Liquid drop out




(c) Gas Z factor

Figure 6.3 continued




(d) Two phase compressibility factor

Figure 6.3 continued

## 6.2. ACCELERATION OF THE PROCESS

Compositional simulation is usually slower than the black-oil simulation. In order to reduce the time for each simulation and the number of simulations, we introduce some of the ideas and technics.

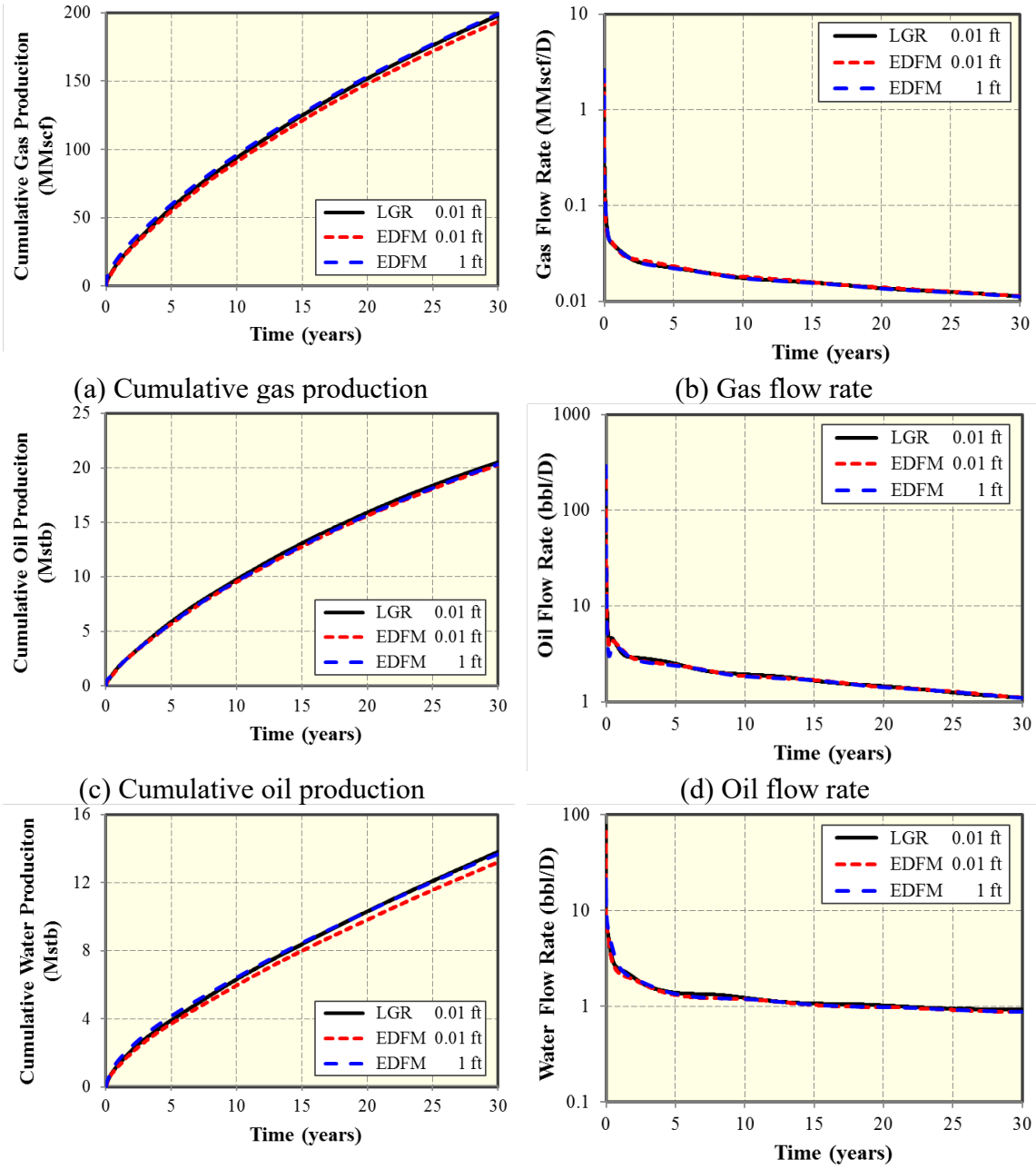
### 6.2.1 EDFM

Although EDFM has been used in the compositional simulator in previous studies, we want to make sure it works on this specific field case. Here we verify the EDFM with traditional local grid refinement (LGR) on a test case that has the same fluid model and petrophysics properties as our target. All the properties are identical except for the fracture modeling technique. The LGR approach refines one parental cell where fractures located into 7 cells parallel to the fracture. Our EDFM preprocessor is used to calculate non-neighboring connections and edit simulator data files according to the fracture descriptions and base matrix grid. The production profile from both methods is compared in Figure 6.3. No significant difference between EDFM and LGR is shown from this case, but we can

see that the CPU computational time of LGR is higher than that of EDFM as listed in Table 6.5. The difference in the computation time increases when we have more fractures in the reservoir. We also tested the model with the fracture width of 1 ft, and adjust fracture permeability to have the same fracture conductivity (Rubin, 2010). The result is then compared with that from cases with fracture width of 0.01. Again, no significant difference is noticed. To conclude, we can perform assisted history matching faster for the unconventional reservoir by using EDFM with satisfiable accuracy.

Table 6.6: Summary of the simulation time for different fracture models

| **Case** | **LGR 0.01ft** | **EDFM 0.01ft** | **EDFM 0.1ft** |
|---|---|---|---|
| **CPU times (s)** | 951 | 842 | 780 |




(a) Cumulative gas production

(b) Gas flow rate

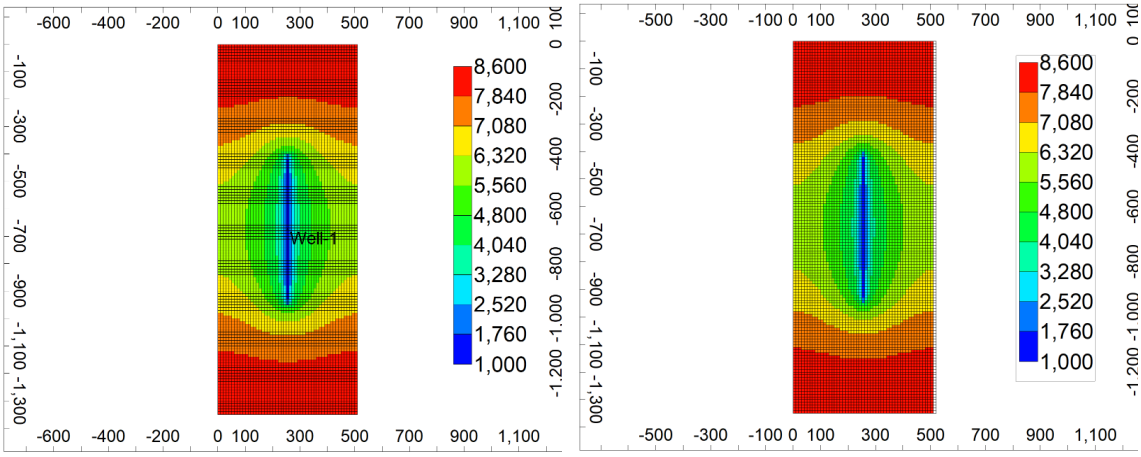
(c) Cumulative oil production

(d) Oil flow rate

(e) Cumulative water production

(f) Water flow rate

Figure 6.4: Comparing the simulated production profile of LGR and EDFM using target reservoir model.




(a) LGR (b) EDFM

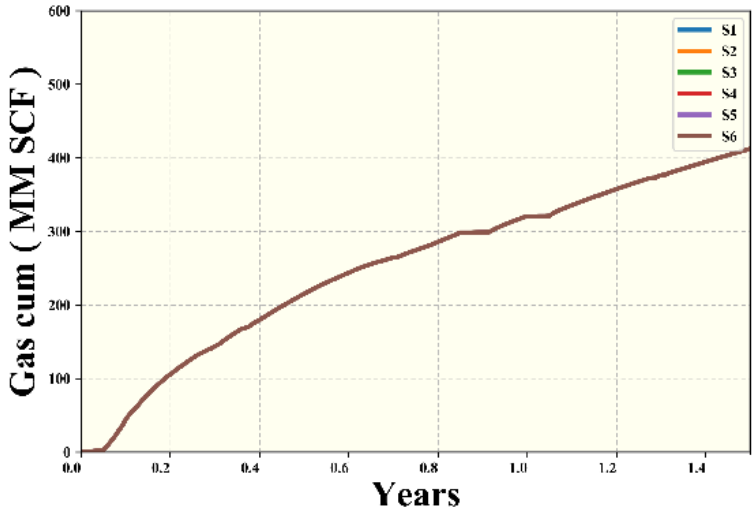
Figure 6.5: Comparing the simulated pressure depletion after 30 years production of (A) LGR and (b) EDFM using target reservoir model.

### 6.2.2 SECTOR MODEL

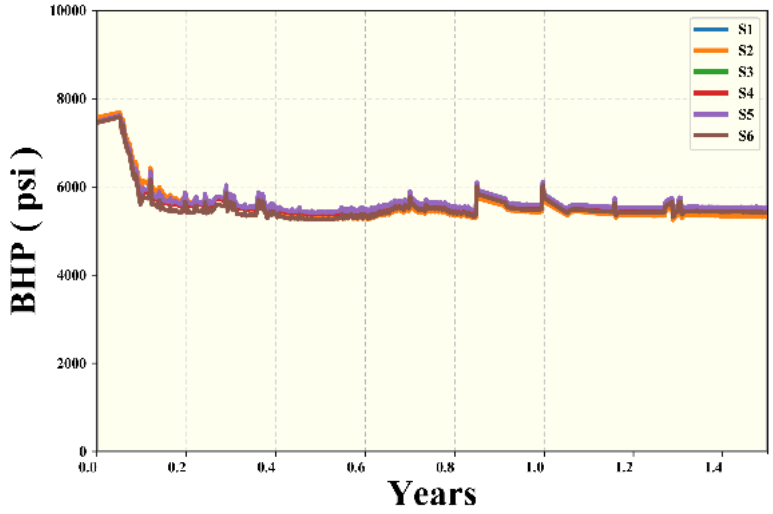
Sector model is used to reduce the computation time further. Since the hydraulic fractures are assumed to distribute evenly along the well, each part around one hydraulic fracture are alike, and the major difference comes from the distance to boundaries. If the effect of the boundary is not significant during the time of interest, we can do the simulation on the part of the reservoir and convert correspondingly. Simulation on the full model and sector models with different numbers of fractures are tested as shown in Table 6.6.

Table 6.7: Sector models accuracy and computational time

| Case Name | Nx | Ny | Nz | Number of Fractures | Accuracy | CPU Time (s) |
|---|---|---|---|---|---|---|
| **S1** | 613 | 201 | 13 | 22 | - | 130,551 |
| **S2** | 307 | 101 | 9 | 22 | low | 9,898 |
| **S3** | 613 | 201 | 13 | 4 | high | 93,328 |
| **S4** | 613 | 201 | 13 | 3 | high | 87,502 |
| **S5** | 613 | 201 | 13 | 2 | high | 58,197 |
| **S6** | 613 | 201 | 13 | 1 | medium | 30,414 |

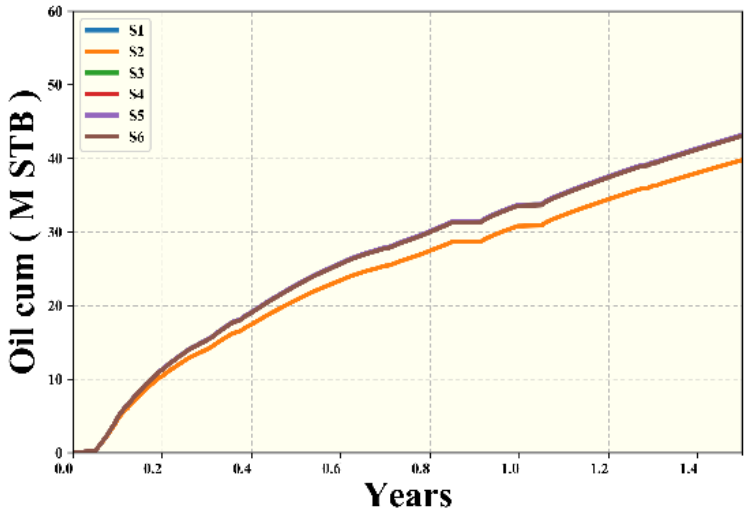



(a) Cumulative gas production

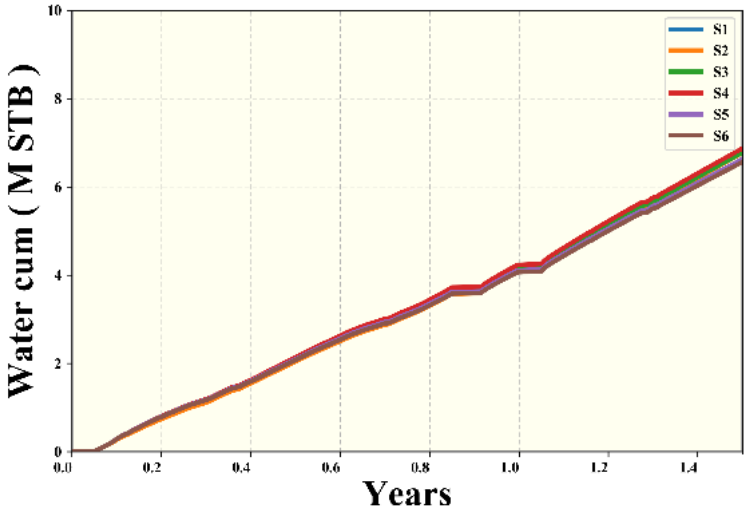



(b) Bottomhole pressure

Figure 6.6: Production profiles that are identical to the benchmark.




(c) Cumulative oil production




(d) Cumulative water production

Figure 6.6 continued.

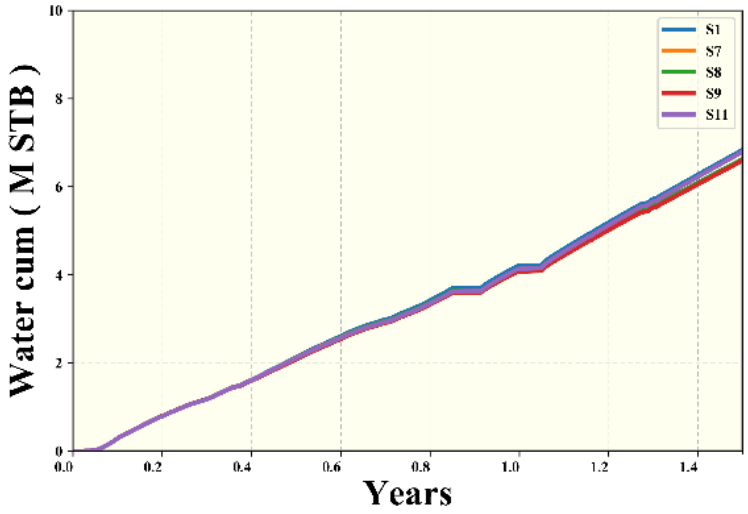
The production profiles of different grids are compared in Figure 6.5. It turns out that the difference between sector models with more than two fracture and the full-sized model is less than 1%, while the one fracture model will have slight discrepancies on BHP. Sector models with two hydraulic fractures are used in the following studies.

### 6.2.3 Grid Size

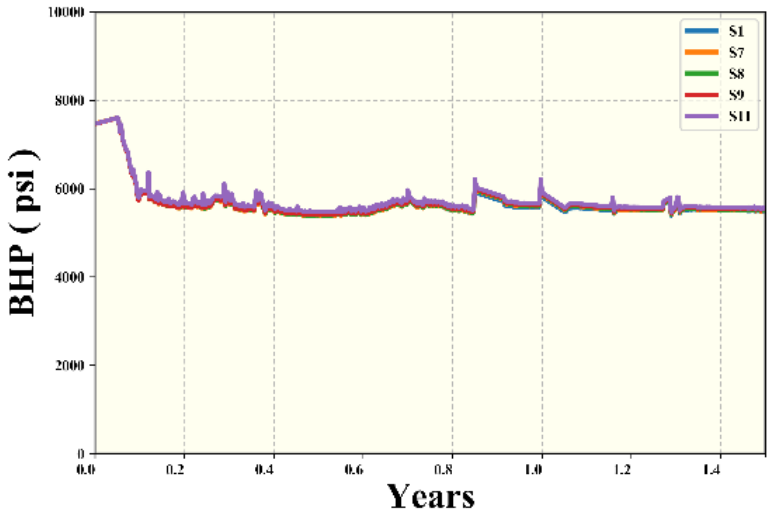
To capture the complex phase behavior usually requires fine grids, yet it also increases the time to do the simulation. Thus we test the effect when we increase the size of grids until the accuracy is compromised. The number of grids in each direction is gradually reduced, and the production profiles are compared with the fine grid full-size model. The tested scenarios of different grid size are listed in Table 6.6. The results that have production profiles of high similarity to the original one are plotted in Figure 6.6, and those are different from our benchmark are plotted in Figure 6.7. The grid size for S11 is used for a later study.

Table 6.8: Grid size and sector model tested

| Case Name | Nx | Ny | Nz | Number of Fractures | Accuracy | CPU Time (s) |
|---|---|---|---|---|---|---|
| **S1** | 613 | 201 | 13 | 22 | - | 130,551 |
| **S7** | 613 | 101 | 13 | 2 | high | 28,296 |
| **S8** | 613 | 51 | 13 | 2 | high | 16,307 |
| **S9** | 613 | 41 | 13 | 2 | high | 13,531 |
| **S10** | 613 | 21 | 13 | 2 | low | 9,037 |
| **S11** | 613 | 41 | 5 | 2 | medium | 4,932 |
| **S12** | 613 | 41 | 3 | 2 | low | 3,155 |

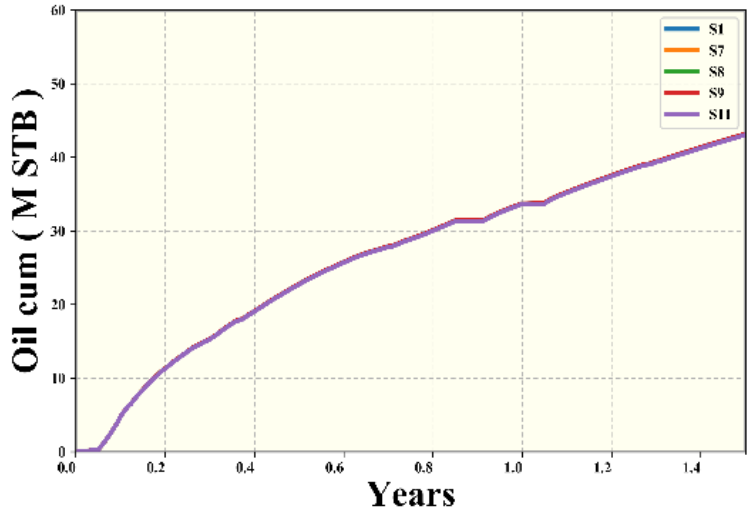



(a) Cumulative water production

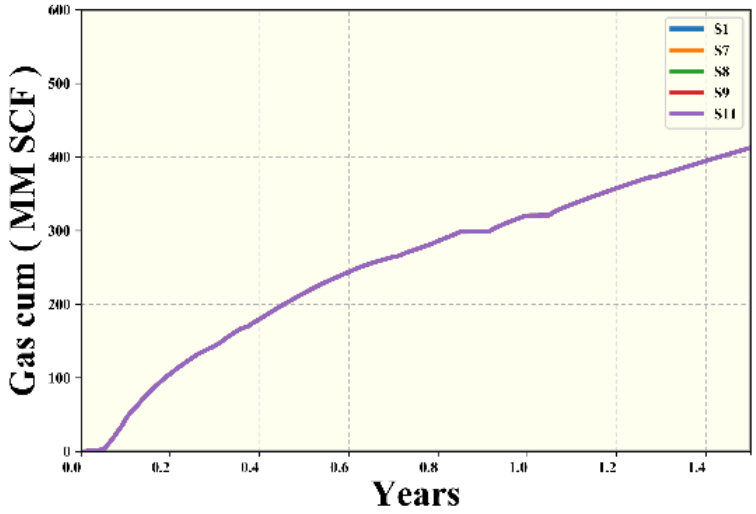



(b) Bottomhole pressure

Figure 6.7: Production profiles that are identical to the benchmark.

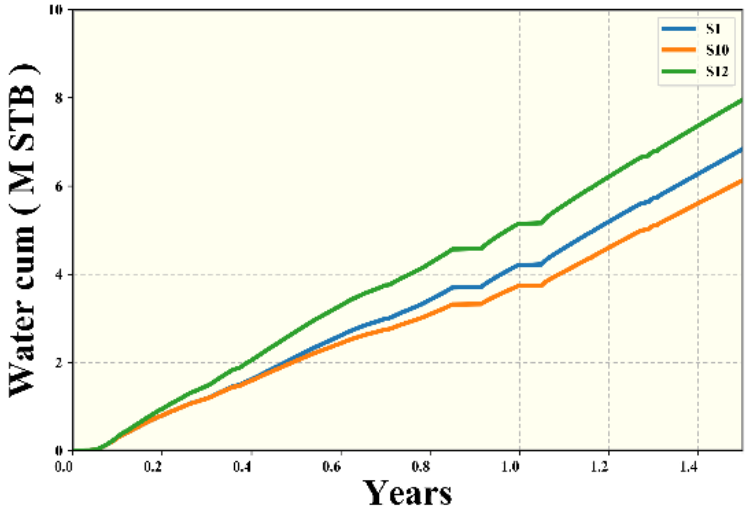



(c) Cumulative oil production

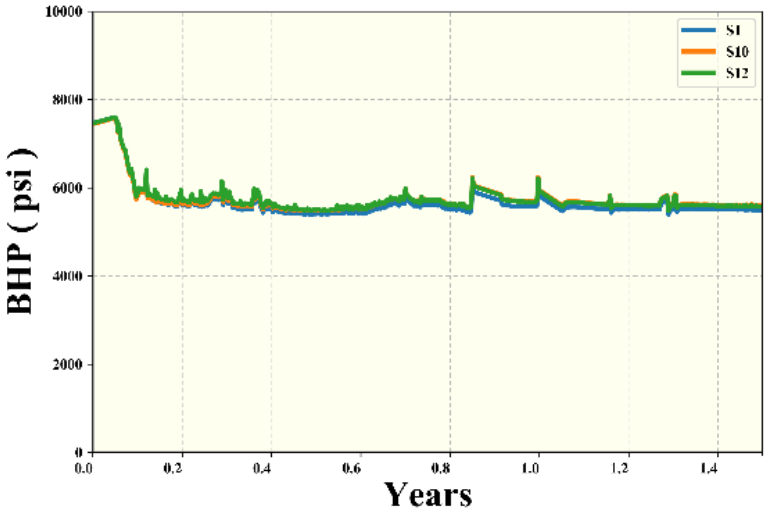



(d) Cumulative gas production

Figure 6.7 continued

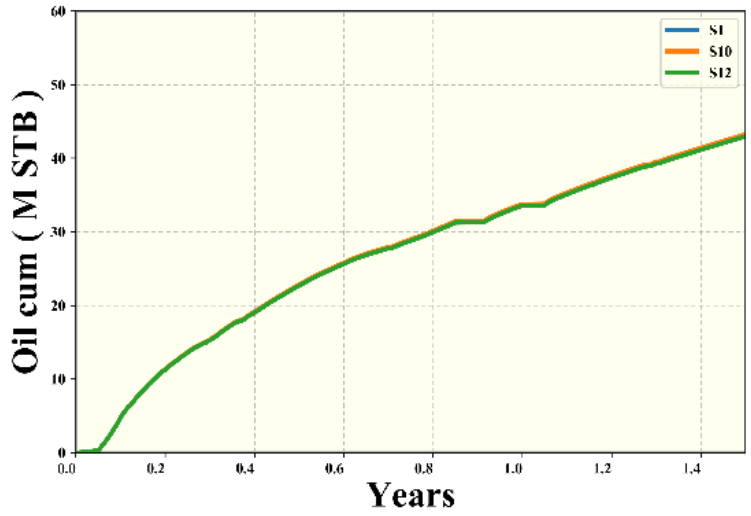



(a) Cumulative water production

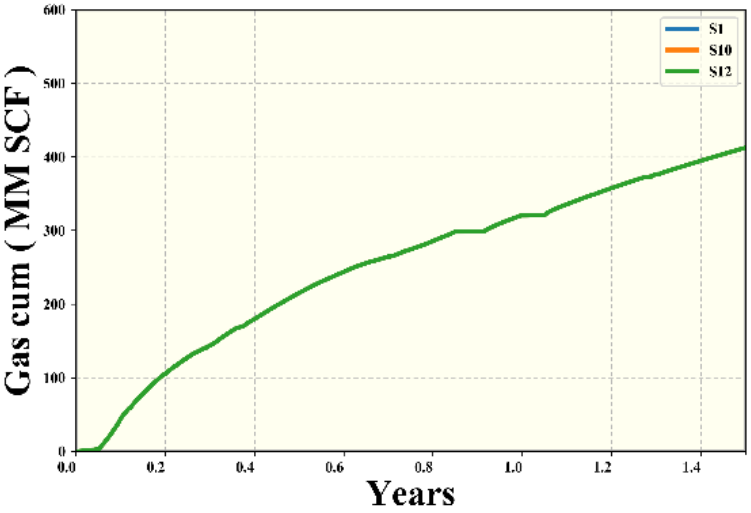



(b) Bottomhole pressure

Figure 6.8: Production profiles that are different from the benchmark.




(c) Cumulative oil production



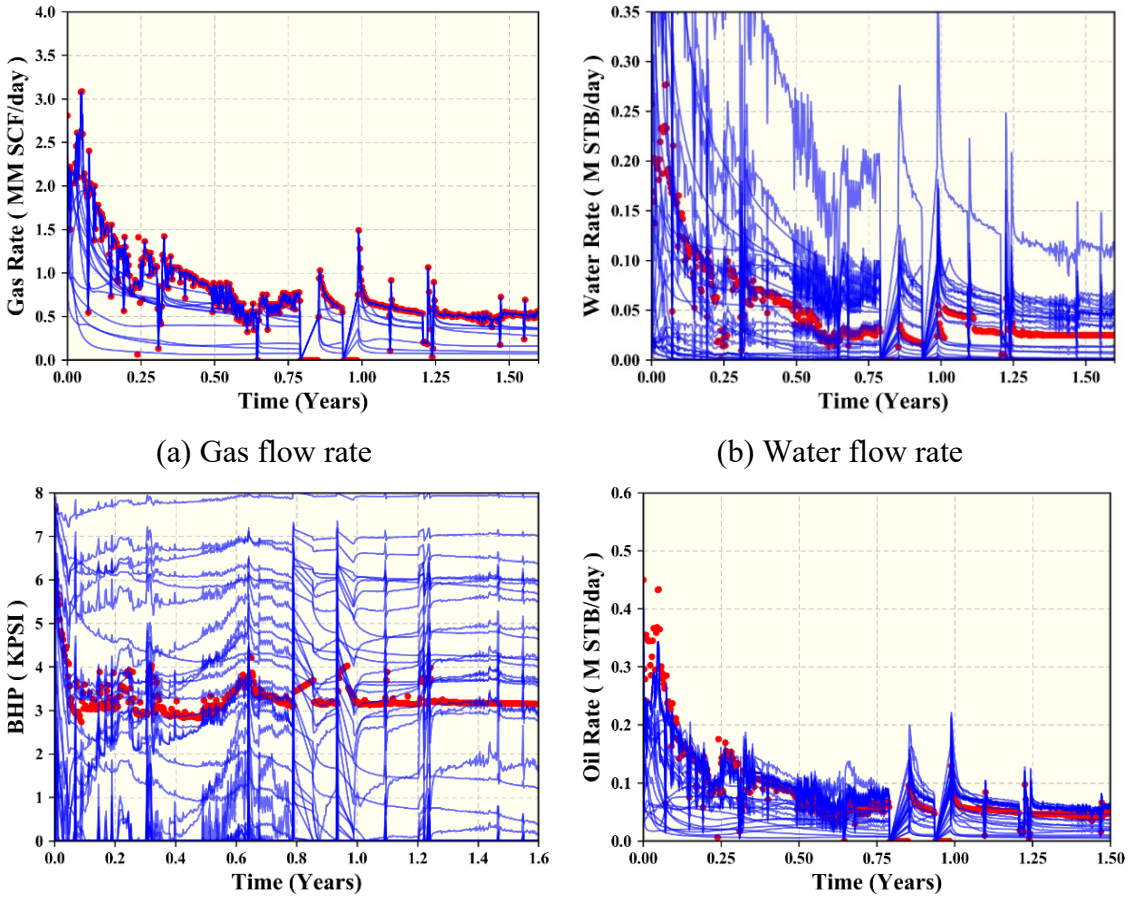

(d) Cumulative gas production

Figure 6.8 continued

### 6.2.4 Initial design and sensitivity analysis

We perform a sensitivity analysis to narrow down the uncertainties to be considered. Only uncertain properties that are statistically important are used in later study. A two-level factorial design is used, and the scenarios designed are evaluated with the simulator. The production profiles of the designed scenarios are compared with the production history in Figure 6.8. The simulated production represented by the blue lines

covers the range of real production history represented by the red dots, which also means an acceptable range of the uncertain properties.




(a) Gas flow rate (b) Water flow rate

(a) Bottomhole pressure (b) Oil flow rate

Figure 6.9: Simulation results of sensitivity analysis and production history.

Analysis of variance is performed to select the significant modeling terms for RMSE. The half-normal plot presented in Figure 6.9 is a popular tool for this purpose. The variable and interaction of variables at the tails are considered to have large effects while modeling. For this study, we want to find the variables that are important for every response rather than the weighted RMSE which represents the overall performance. It will be better

if the variance is studied directly. We can decompose the variance of each response to different variable and their interactions.

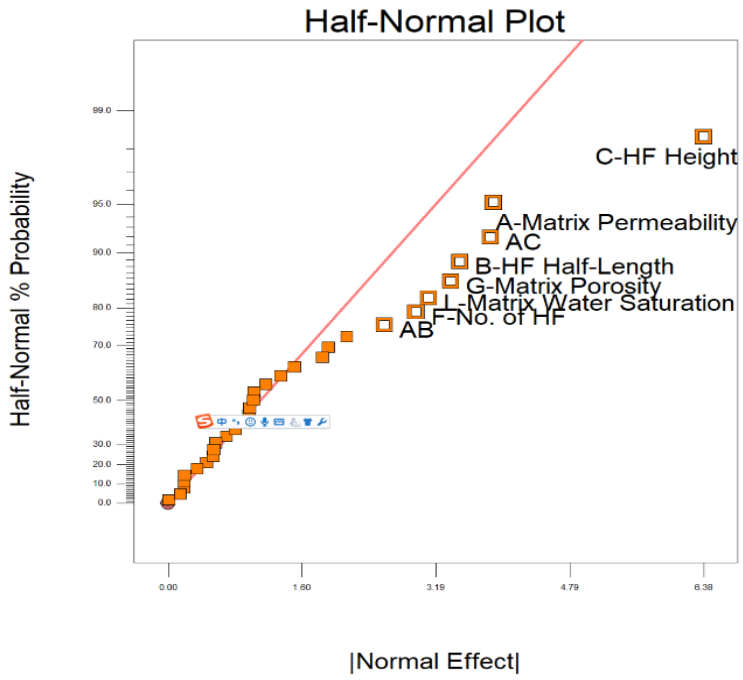



Figure 6.10: Half-normal plot of the weighted RMSE.

To exclude the effects of the scale, the sum of variance for each response is normalized to 100. A larger value represents greater importance when a linear model is built. The variance of oil rate gas rate, BHP, and weighted RMSE are decomposed in Table 6.7. Each column represents a response variable of interest, and each row represents an uncertain variable to be studied. The sum of the variance caused by single properties are calculated in the table and are not equal to 100. Those part can be explained when the interactions of several properties are added. Notable interactions are also listed at the end of the table. Both properties that have large influence individually or together with other properties are kept in the further study and marked as 1 in the last column of the table. One thing worth mention is that this kind of analysis of variance is based on linear assumption and is only acceptable for parameter screening. The variance shown here should be only regarded as a reference as we are going to build models that are not linear. 9 uncertain

variables are regarded as important and kept for the next step, other uncertain properties to be the average of its boundary. The uncertain parameters are reduced to six parameters while the other four parameters become constant.

Table 6.9: Contribution of each input to each response

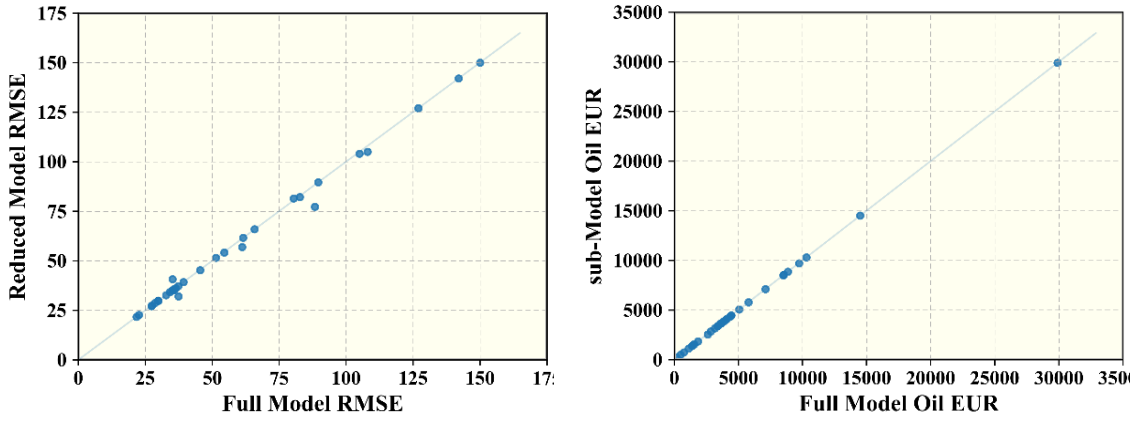
| | Oil Rate | Gas Rate | Pressure | Sum of Error | Use |
|---|---|---|---|---|---|
| **A-Matrix Permeability** | 12.81 | 8.5704 | 6.3010 | 10.590 | 1 |
| **B-HF Half-Length** | 9.747 | 7.6556 | 1.1173 | 8.5180 | 1 |
| **C-HF Height** | 25.36 | 21.90 | 25.983 | 28.685 | 1 |
| **D-HF Conductivity** | 0.0153 | 0.6578 | 0.6412 | 0.0275 | 0 |
| **E-Equivalent Fracture length** | 0.0214 | 0.8270 | 5.7659 | 0.1551 | 1 |
| **F-No. of HF** | 4.328 | 8.2141 | 0.9581 | 6.1495 | 1 |
| **G-Matrix Porosity** | 12.10 | 6.0588 | 2.3420 | 7.9846 | 1 |
| **H-Fracture Water Saturation** | 6.317 | 0.3147 | 12.040 | 1.5170 | 1 |
| **J-Langmuir Inverse Pressure** | 0.2757 | 2.1545 | 0.0184 | 0.3503 | 0 |
| **K-Rock Compressibility** | 0.7706 | 0.1619 | 1.0869 | 0.2365 | 0 |
| **L-Matrix Water Saturation** | 13.98 | 6.8599 | 23.125 | 6.7989 | 1 |
| **M-Langmuir Volume** | 0.316 | 2.1470 | 0.0979 | 0.0001 | 0 |
| **sum of single item** | 86.05 | 65.529 | 79.479 | 71.014 | |
| **Notable interaction** | AF, AG, AE | AB, AF, AG, AE | AB, CM, EJ | AC, AB, AM | |

Table 6.10: Summary of the uncertain parameters after the parameter reduction.

| Type | Parameter | Unit | Min | Max |
|---|---|---|---|---|
| Uncertain | Matrix permeability | mD | 0.00002 | 0.0004 |
| | Hydraulic fracture half-length | ft | 100 | 700 |
| | Hydraulic fracture height | ft | 20 | 100 |
| | Equivalent fracture length | ft | 0.01 | 0.3 |
| | No. of effective hydraulic fracture | - | 13 | 27 |
| | Matrix porosity | - | 0.02 | 0.15 |
| | Fracture water saturation | - | 0.05 | 0.95 |
| | Matrix water saturation | - | 0.1 | 0.8 |
| Constant | Langmuir volume | gmol/lb | 0.056 | |
| | Langmuir inverse pressure | 1/psi | 0.001 | |
| | Fracture conductivity | md-ft | 10 | |
| | Rock compressibility | 1/psi | 0.0000055 | |

### 6.2.5 Validation with more scenarios

To make sure that the consistency of the original model and accelerated model is not scenario dependent for this reservoir, we close the logic loop by comparing more scenarios. All the scenarios used in sensitivity analysis run with the full-size model and compared with the sub-model we are to use in the history matching process. As shown in Figure 6.10, the EURs are identical while there are slight discrepancies for RMSE. We regard the model acceptable as the time of each simulation is reduced form more than two days to about twenty minutes at an ignorable cost of accuracy.




(a) Comparison of RMSE (b) Comparison of oil EUR

Figure 6.11: Comparison of the multiple-fractured full model and sector model.

### 6.3 Sampling Strategy for Solution Diversity

The history matching method to be used in this chapter is based on the iterative response surface method. The scenarios chosen to be evaluated by the simulator at each iteration will affect the performance of this method. As discussed in Section 3.2.1.5 and Section 3.2.2, well-designed sampling strategy can help us achieve diversified solutions.

In this study, a sequence of four samplings units are used to select scenarios for simulation and the workflow is shown in Figure 6.11. We have an initial sampling unit to generate a large number of points, and then the objective function, weighted RMSE, is evaluated by the proxy model. Those points are then fed into an optimum unit to do the first screening, only points that have not too large errors are kept and fed to the next diverging unit. This unit will now get points that are remote from each other and send to the next unit. This unit again will find optimum points and send to the simulator.

With four units above we can limit the distribution pattern of the sample we use. Also, by setting the sample output size for each layer, we can switch the goal. For the first optimum unit, more samples lead to more diverse results while fewer results in more accurate solutions. For the diverging unit, more output results in less diversified samples while fewer output results in more diversified results.

We choose the governing equation to decide the number of outputs for each unit and thus the ultimate sampling pattern. While a more complex set of rules coupled with the judgment of the prediction may improve the accuracy, a set of equations are chosen to decide for a wanted pattern. As a limited number of simulation is planned, we want the workflow to do more exploration on early stage and concentrate more on optimizing around solutions in the later stage. Thus, we decrease the output of the first optimum unit and increase the output of the diverging unit. The numbers of samples in the candidate pool at different iteration are shown in Figure 6.12.

The purposed method in this study is not focusing on giving universal rules to search the optimum but trace back to the first principals behind and propose a way to balance them during the searching process. We suggest that the sampling method should be tuned base on the objective and operating limits.

Design a sampling algorithm to satisfy multiple purposes may not be easy. However, just like we can decompose one massive problem into several small ones, we can reach our sampling goals by utilizing different sampling methods together.

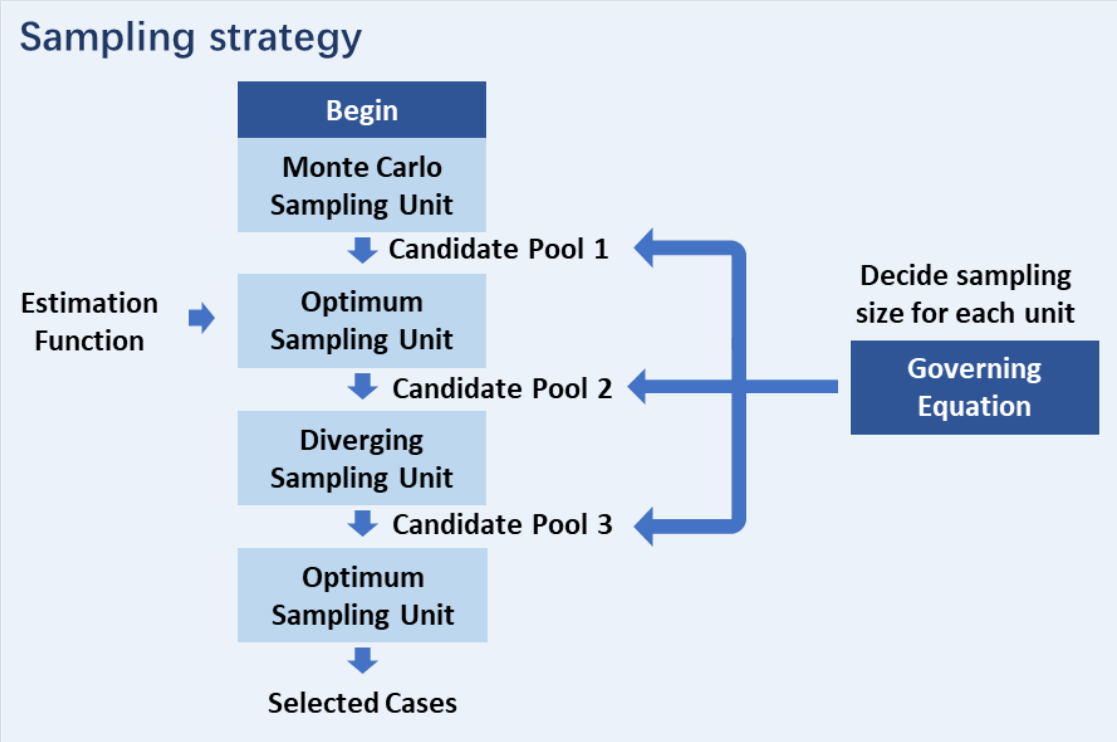



Figure 6.12: Sampling units used in this workflow.

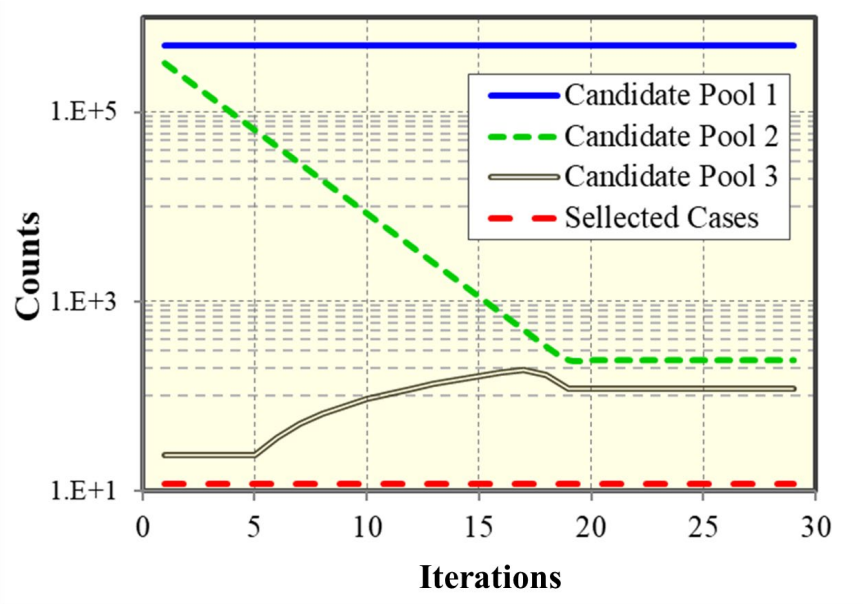



Figure 6.13: The effect that controls the governing equations in this study.

## 6.4 HISTORY MATCHING

The workflow in Section 3.6.2 is used for the assisted history matching.

### 6.4.1 Process

The results of 32 cases we run during parameters screening phase are used directly as the first batch of training dataset for the ANN. Then the ANN is used as a proxy model that helps estimate potential cases for simulation and will be updated after the simulations of each iteration until the stopping criteria are reached. The 12 scenarios selected at each iteration is not only based on how well the proxy predicts it to be but also the similarity to already-exist scenarios.

The proxy model is implemented base on the workflow in Chapter 4 since we have very few information about the nonlinearity of reality, especially at the early stage. Figure 6.13 illustrates the improvement of proxy predictability in weighted RMSE. We compare the predictions of the proxy and the simulation results for the newly added batch of scenarios. The points are more concentrating on the X=Y lines as iterations go on, which indicates the improvement of the proxy's predictability for newly dataset. We can also notice that the model is not good enough for accurate prediction of the exact value. This acceptable, as the primary goal for the proxy model here, is merely assist in designing the scenarios.

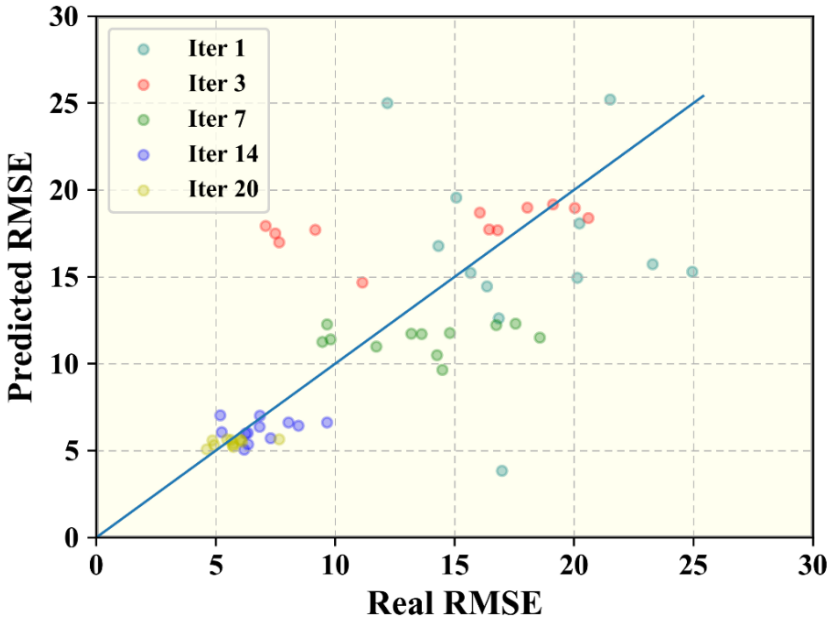



Figure 6.14: The proxy predicted RMSE vs. simulation RMSE in different iterations.

The accuracy of solutions will be improved as iteration progresses, but for practical application, we need to stop when we have acceptable results. One may assume a maximum and minimum amount simulations that are planned to run. Besides that, we raise another two principles to terminate the workflow, error reduction and proxy convergence.

Error reduction criteria is a more obvious principle that we stop further testing when the solutions do not improve. Proxy convergence shows that the proxy understands the problem and does not change much.

The error reduction criteria compare the error from the new batch of simulation points and previous results. Since the objective function is a weighted average of multiple RMSE, and the base model may not be comprehensive enough to capture the heterogeneity of the reservoir, it is not plausible to set a threshold to decide if the accuracy cannot be improved. When the best point from the new batch is not improving by 5% and the average

of the new batch is no longer better than both the previous two batches, we think the results are not improving.

The convergence of proxy is tested by comparing the predictions of the current proxy and that from the previous step on the same set of points. If the difference is within a specific threshold, the proxy is regarded as stable. In the field case example, 5% is used, and 100,000 random points are chosen. As shown in Figure 6.14, the difference of the prediction of each iteration with its previous iteration on 100,000 scenarios is more closer to each other. Although it is partly caused by the fact that the number of newly added points, which is the primary source of difference, will be less compared with several already existing points as iteration goes on, it still provides a noticeable hint on when the convergence is approaching.

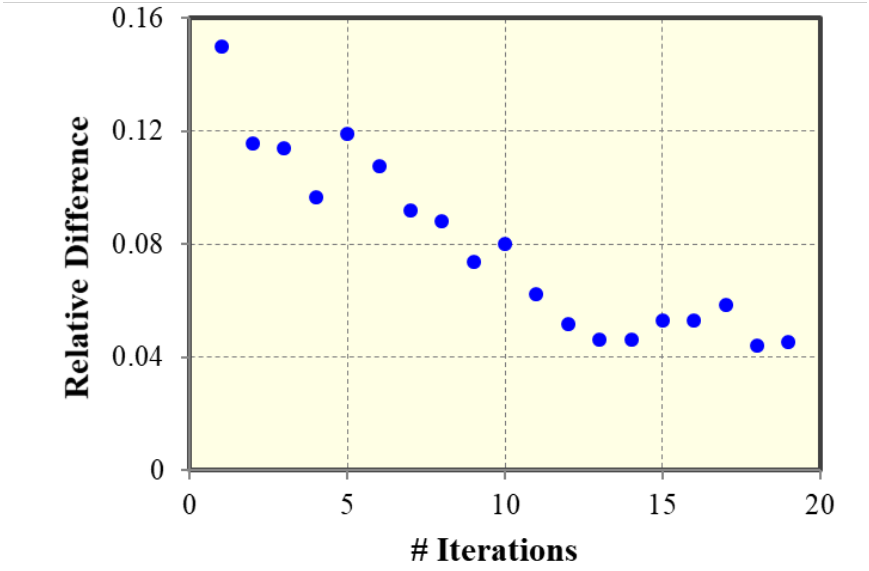
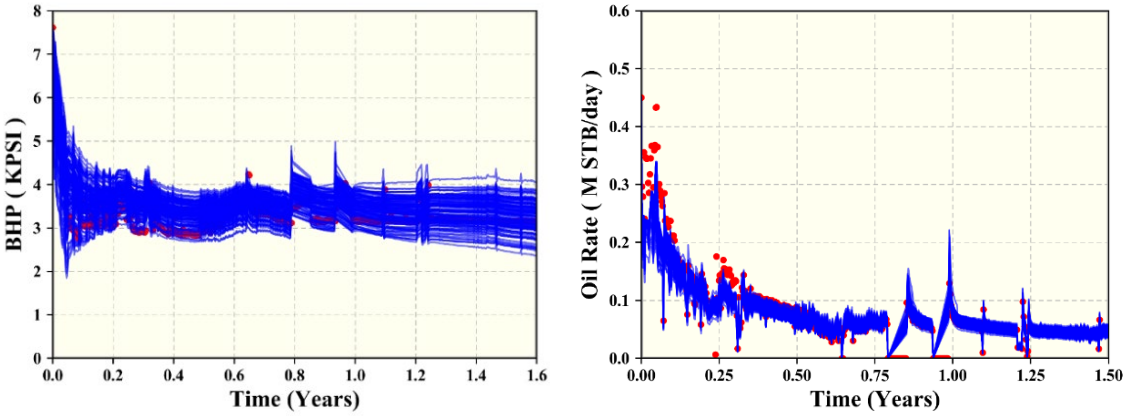



Figure 6.15: Relative differences between the prediction of the proxy model of current and previous iteration on a large set of scenarios.

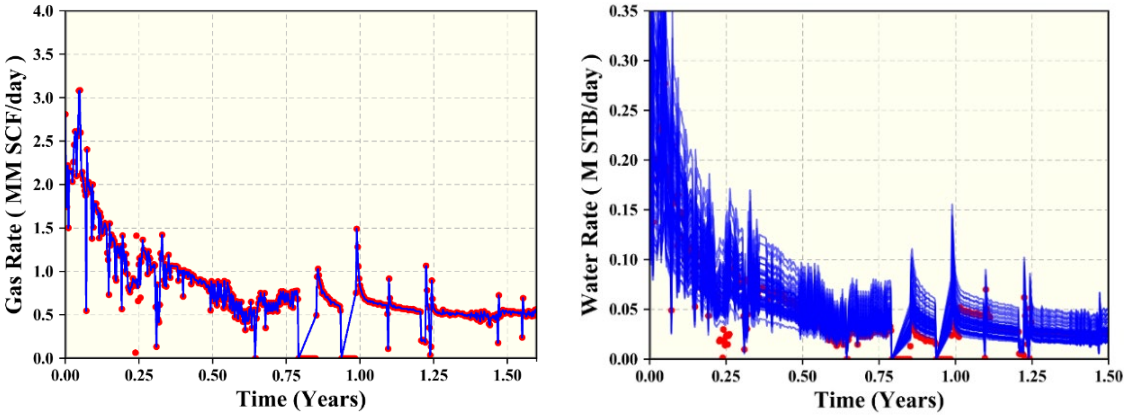
### 6.4.2 History Matching Solutions

In this study, the workflow stops at iteration 20. 240 simulations are run, and 140 are regarded as the solutions of history matching. The production profiles of all the history matching solutions we choose are presented in Figure 6.15. The improvements in HM

solutions are shown in Figure 6.16. The best solution of each batch is improving, and the chosen simulation points are also concentrating on lower error region as iteration goes on. The points that have not such good performance is because we planned those points to work on regions that have not been explored.




(a) Bottomhole pressure (b) Oil flow rate




(a) Gas flow rate (b) Water flow rate

Figure 6.16: 50 history matching solutions chosen from the 272 simulations.

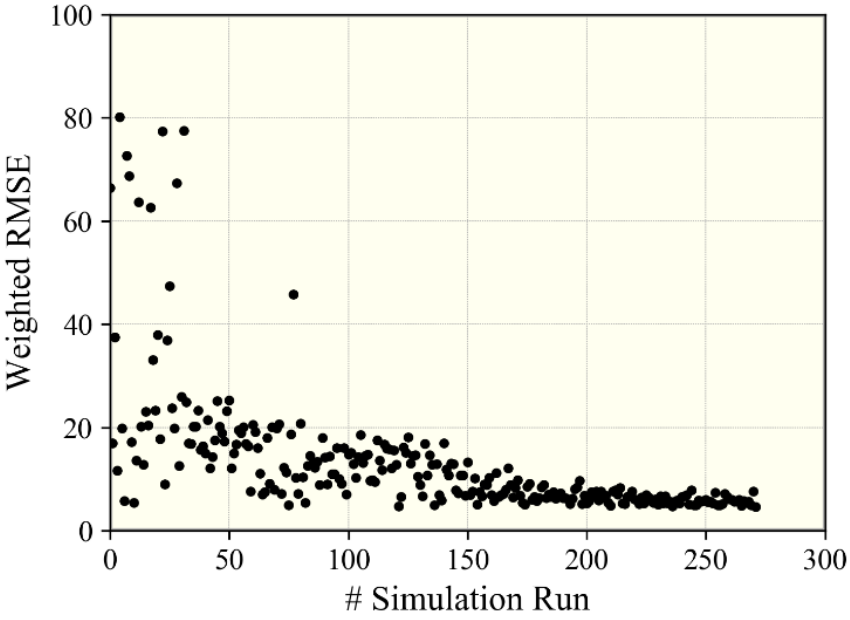
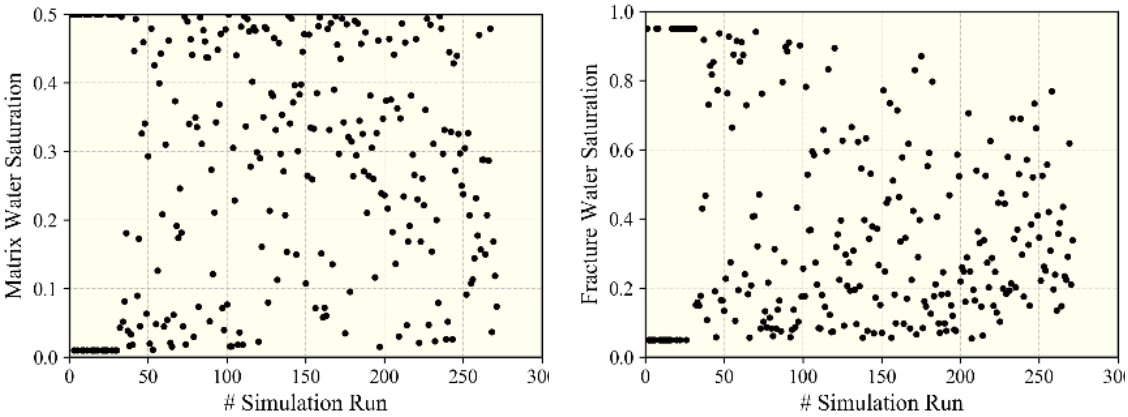



Figure 6.17: Weighted RMSE in the process of AHM workflow.

### 6.4.3 Effect of Diversity Control

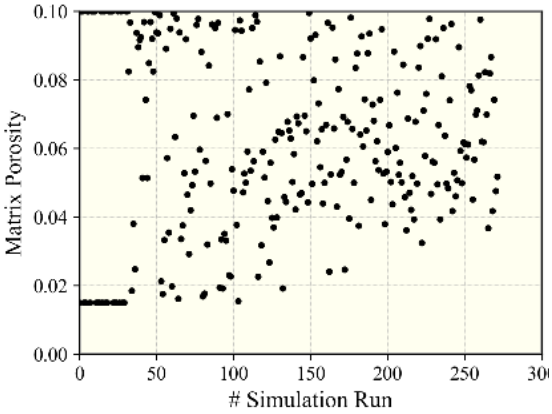
Solution diversity is the main driving force behind this workflow, and we can see from Figure 6.17 how the solutions are distributed and how many regions we have explored and confirmed the non-existence of the solution. The effect of the governing equation can also be directly seen from the distribution of the uncertain variables throughout the workflow. We can see that the algorithm is truly exploring a larger area and then concentrating on the specific part at a later stage.



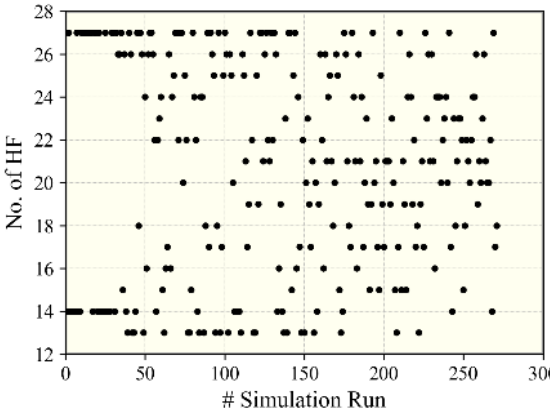

(a) Water saturation of the reservoir matrix



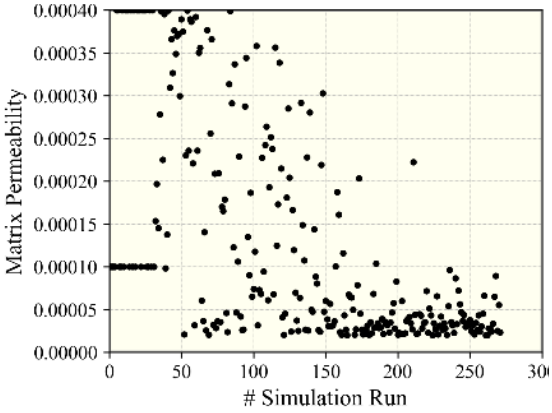

(b) Water saturation of the hydraulic fracture



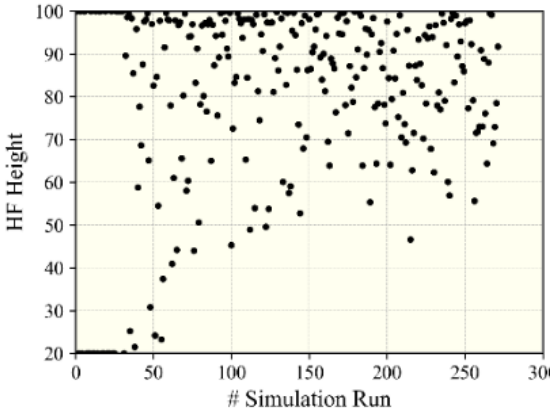

(c) Porosity of the reservoir matrix




(d) Number of the hydraulic fracture

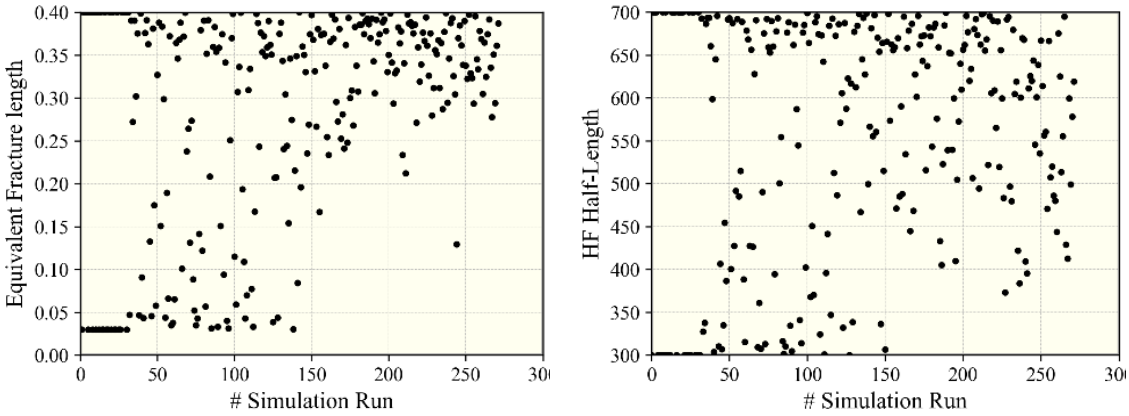



(e) Permeability of the reservoir matrix




(f) Height of the hydraulic fracture

Figure 6.18: Uncertain variables searched by AHM algorithm.




(g) Equivalent width of the hydraulic fracture

(h) Half-length of the hydraulic fracture

Figure 6.18 continued

The scenarios that are regarded as solutions are presented by a parallel coordinates plot in Figure 6.18. The values of each uncertain property are normalized according to their range. Each connected line in the plot represents a simulated scenario; the parts with denser lines have more possibility for the solutions and is thus explored more. The region with fewer lines usually have larger errors, and we tend to explore it only once.

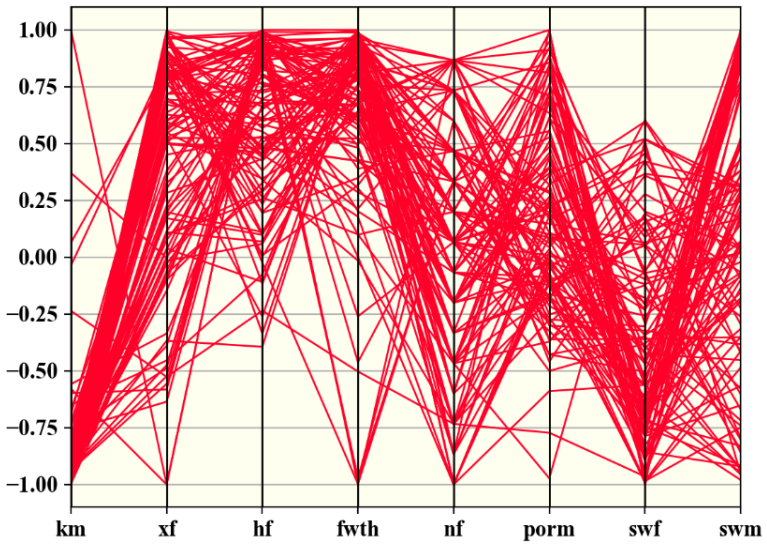



Figure 6.19: Normalized HM solutions presented with the parallel axis plot.

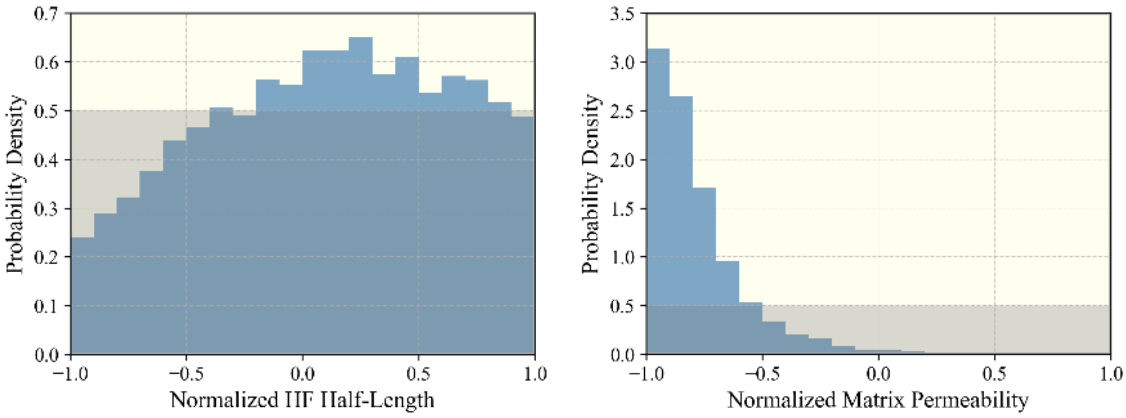
## 6.5. SOLUTION ANALYZATION

The resources needed to do the simulation for this kind of reservoir is so large that doing too many simulations may not be a good idea. Approaches to utilize the results of history matching is introduced here.

### 6.5.1 Uncertainty Quantification

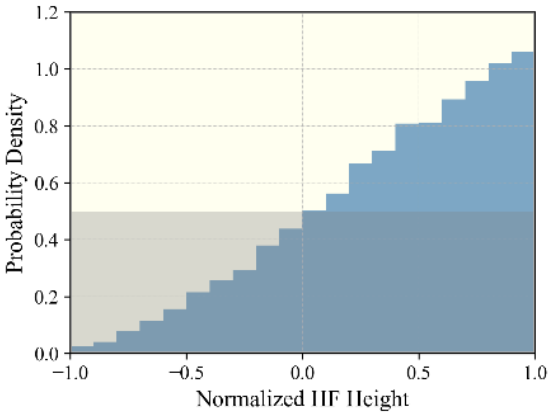
MCMC can be used to generate a real distribution of the uncertainties. Random walk MCMC is used here to estimate the probability of final production and quantify the uncertainties of the uncertain variables. All the 240+32 simulation results are used here to build two proxy models. The error proxy is trained by the weighted objective function and is used to estimate the possibility of a random walk. We used 100,000 cases MCMC chain and applied 20% burn-in. The acceptance rate is 26%, close to the recommended value.

The prior and posterior distribution of each uncertain parameters are shown in Figure 6.19. We notice that for the matrix permeability, matrix water saturation, fracture water saturation, fracture height, equivalent fracture saturation, the posterior distribution have been narrowed to a smaller range. This implies that those parameters are susceptible to the objective function. However, it is not proper to conclude that other parameters are less critical as it is the combination of all the parameters that influence the matching results.

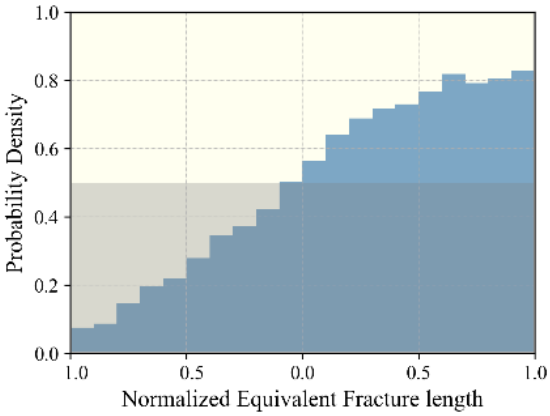



(a) Fracture half-length (b) Matrix permeability

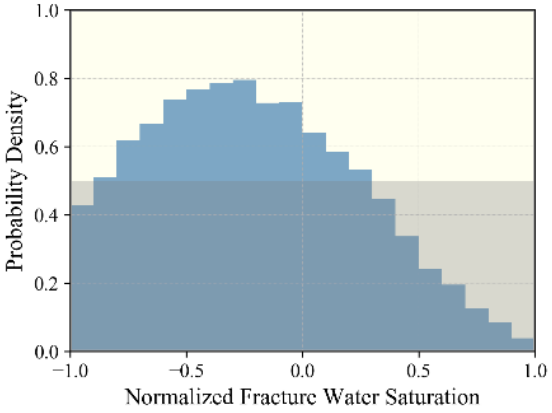
Figure 6.19: Distribution of uncertain parameters after history matching.



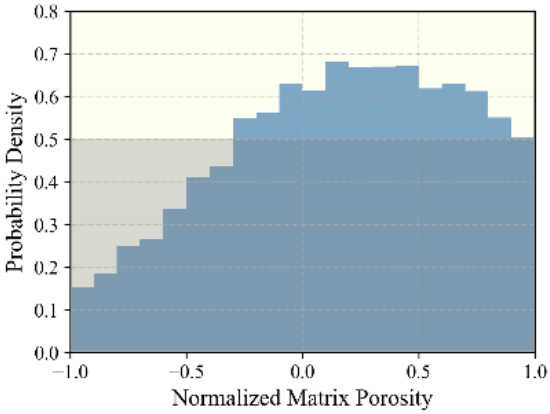

(c) Fracture height



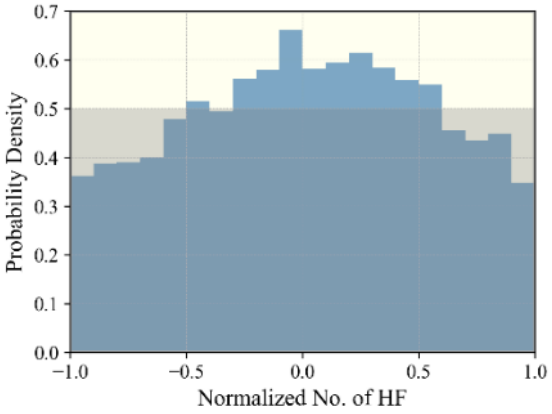

(d) Equivalent fracture length



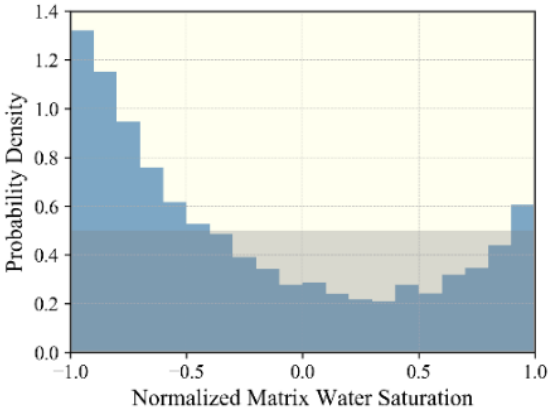

(e) Fracture water saturation




(f) Matrix porosity




(g) Number of hydraulic fractures




(h) Matrix water saturation

Figure 6.19 continued

The parallel coordinates plot can help us understand the relations of different properties of the MCMC solutions. We judge whether a scenario is a solution by its PROXY predicted RMSE, and again 10 is used as a critical value. In Figure 6.20, all the solutions are normalized and plotted with opacity. The solution with the darker line is closer to reality since it is less likely to walk away from those scenarios during the random walk MCMC sampling process.

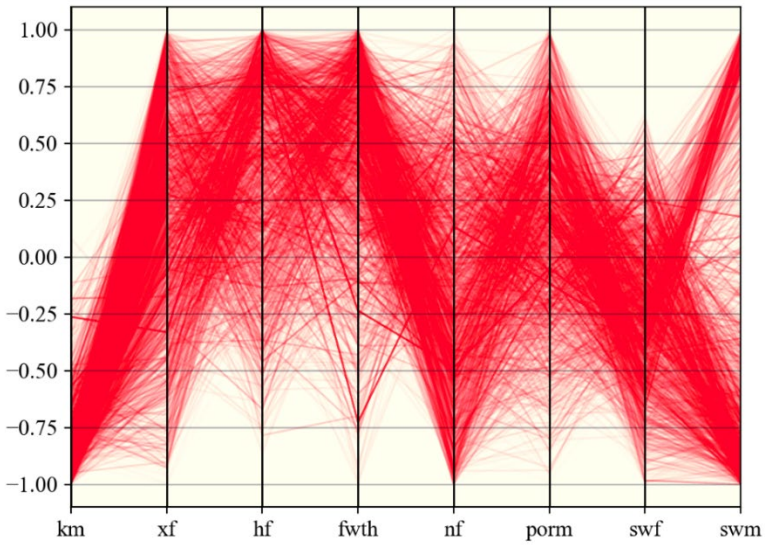
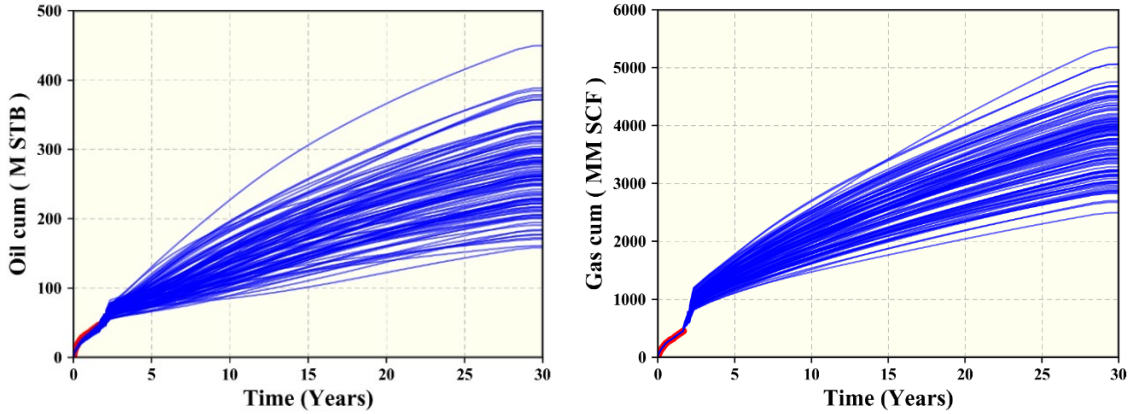



Figure 6.20 Normalized HM solutions presented with the parallel axis plot.

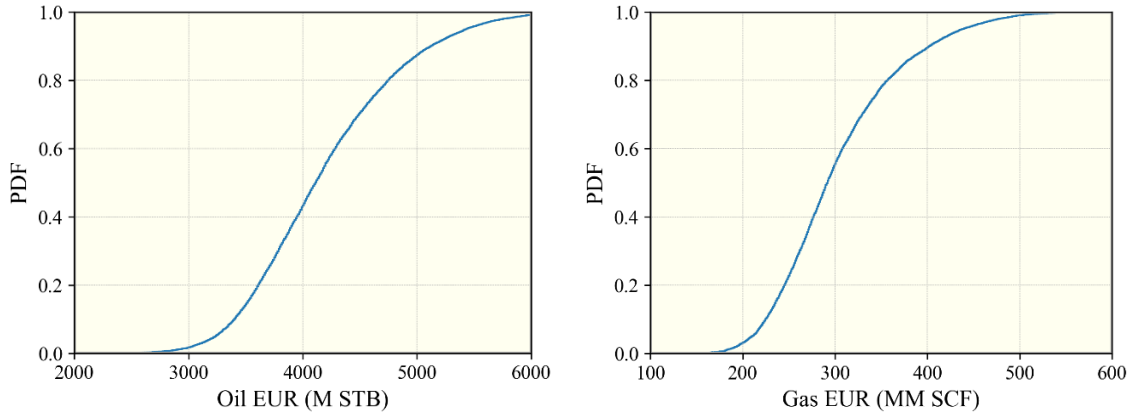
### 6.5.2 Probabilistic Forecasting

The 30-years production profiles of history matched scenarios are shown in Figure 6.20. The EUR proxy is trained and used to estimate the production scenarios generated by MCMC. The empirical cumulative distribution function (ECDF) of gas EUR and oil EUR from all HM solutions are shown in Figure 6.20. P10, P50, and P90 of oil EUR are 3.4, 4.2, and 5.1 MMSTB. The P90/P10 ratio of OIL EUR is 1.50. P10, P50, and P90 of gas EUR are 0.23, 0.28, and 0.40 BSCF. The P90/P10 ratio of gas EUR is 1.74.




(a) Cumulative oil production (b) Cumulative gas production

Figure 6.21: EUR prediction with the HM solutions directly.




(a) Oil EUR (b) Gas EUR

Figure 6.20: PDF of EUR prediction base on the samples drawn from MCMC.

### 6.5.3 Representative Solutions

It is also a practical idea to find representative solutions from all the solutions we get, especially when we want to test the development plans through simulation. In the previous chapter, an example that test different well spacing plan with a bunch of simulations is shown. That is a black oil model, and the simulation for a full-sized model is less than half an hour. For this composition simulation, the simulation for a full-sized

model will take days and placing multiple wells will take weeks; the requirement for memory is also going to be significant.

Clustering is a promising approach to find a representative solution. The workflow is rather simple when well-developed clustering library from Scipy is used here.

1. Set the number of clustering point for the sampling algorithm. While designing a theory to decide how much clustering center can be attractive, in this study only a naïve approach that chooses the number of clustering center base on need is demonstrated.
2. Cluster the normalized history matching solutions. The normalization is necessary since the range and scale of different properties vary a lot.
3. For each clustered group, choose the solution with the lowest weighted RMSE as the representative solution.

For this case, the solutions are clustered to 6 groups, as can be seen in Figure 6.21, the solutions in one plot are alike, and the solution in different plot tend to have a larger difference. We select the solution with the lowest matching error from each group and regard them our simplified solution set. If there are further simulations needed for the development plan and not too much simulation time is available, those solutions can be used as a quick insight into the problem. The simplified solution set is shown in Figure 6.22.

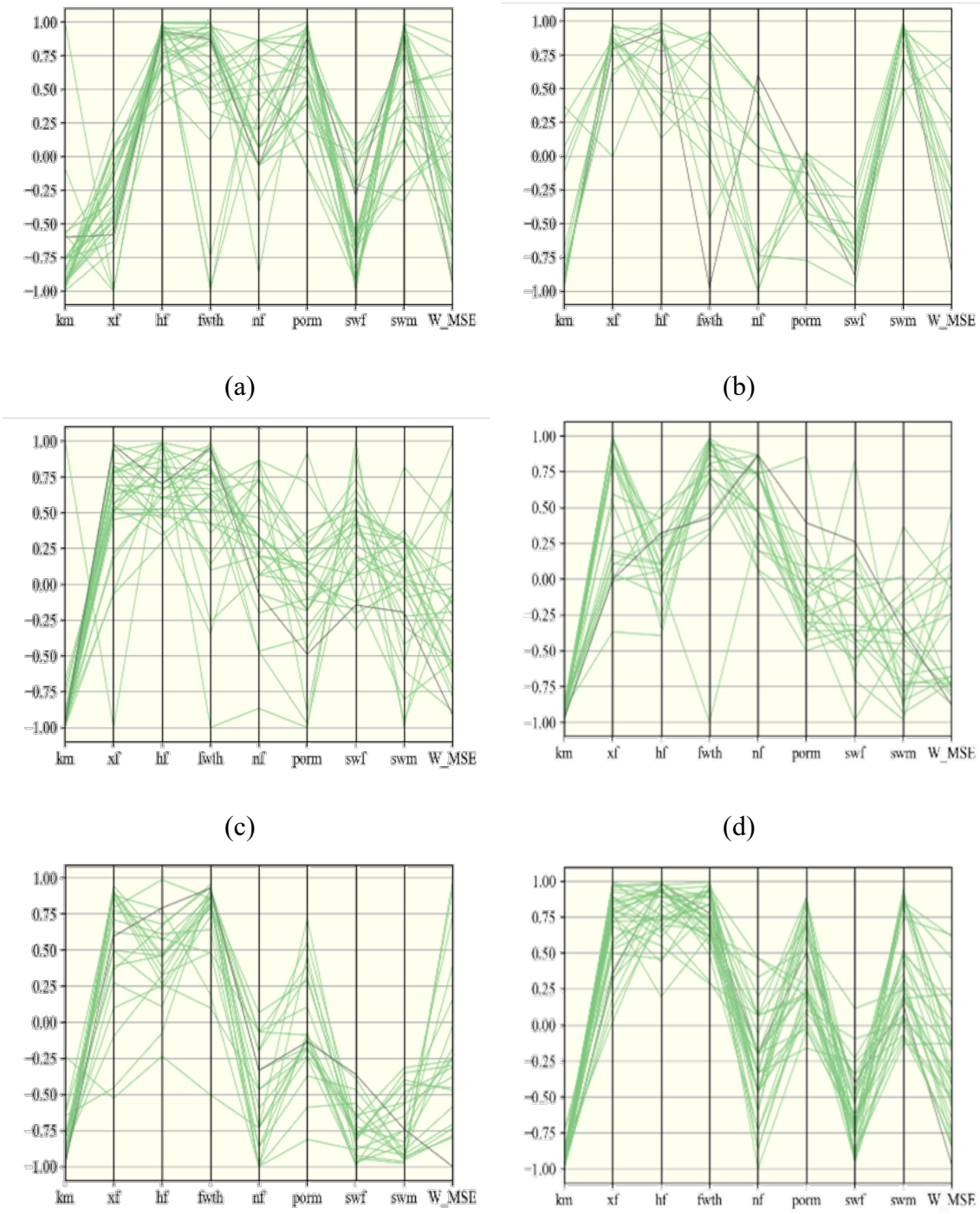



Figure 6.21: History matching solutions clustered into 6 groups.

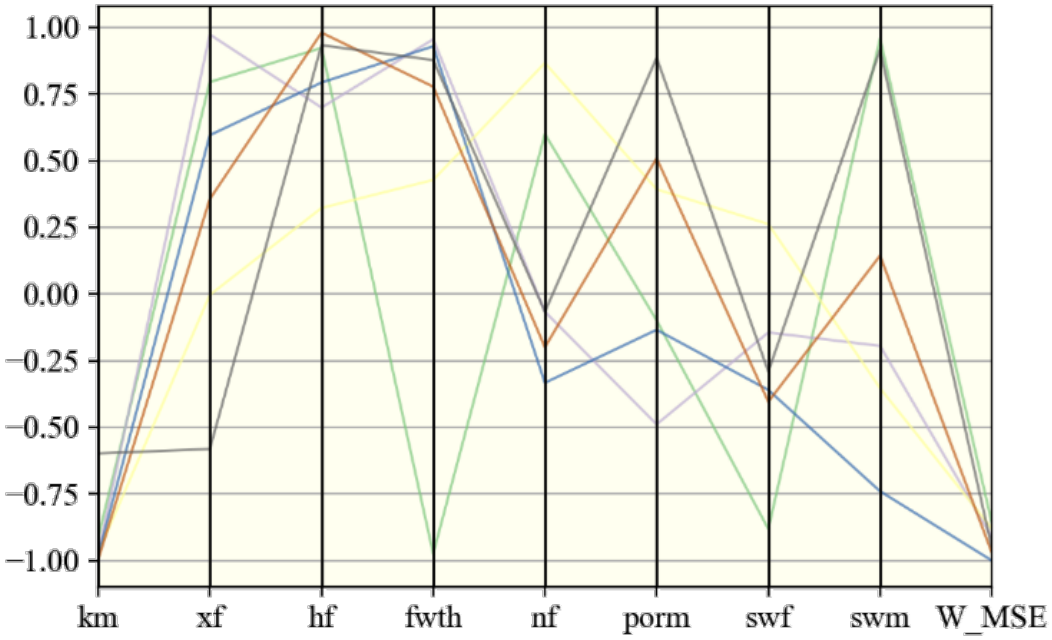



Figure 6.22: Representative solutions when solutions are clustered into 6 groups.

## 6.6 CONCLUSIONS

The assisted history matching workflow in Chapter 3 was applied here on a well in the gas condensate reservoir. The works and findings are summarized as follows:

1. The compositional reservoir model is built for the gas condensate reservoir. Sector model, EDFM, adjustments of the grid size are used to accelerate the simulation.
2. 9 uncertain properties are used for history matching after the screening of sensitivity analysis. The weighted RMSEs of all the production profiles are used as the objective function to be minimized during the workflow.
3. The iterative response surface method is used to perform history matching. ANN is used as response surface and shows acceptable results. The sampling unit and

governing equation are used to implement the designed sampling strategy that controls diversity and performs the optimization.

4. After 20 iterations of history matching, 272 scenarios are evaluated by the simulator, and 110 of them are regarded as history matching solutions according to our criteria.
5. 30 years productions are simulated after history matching for the probability forecast. The results are visualized, and a EUR proxy model is built.
6. MCMC is used to quantify the uncertainties. A sample that can represent the real distribution of uncertain property is generated, and the marginal distribution of the uncertain properties are calculated. The samples are then evaluated by the EUR proxy for probability forecast and provide the P10-50-90 of EUR.
7. The representative solutions of the history matching are then chosen from all the scenarios by clustering in case further simulation work is needed for another purpose.

# Chapter 7: Development of an Interactive Parallel Coordinates Plots

This chapter presents the methods to utilize simulation results of a history matching project with the parallel coordinates plots. The example in this chapter is based on the real field case we history matched in Chapter 6. First of all, the components of the interactive parallel coordinates plot implemented in the platform are introduced. Then, we use it directly on the simulations that have been run to look into the details of the simulation results. After that, we demonstrate the workflow that involves evaluating the scenarios with the proxy model. By designing scenarios randomly, for one reservoir realization and multiple representative reservoir realizations, different example applications are shown. Finally, for this game-changing way to visualize higher dimensional data and model, the potentials and limitations are discussed.

## 7.1 Digging Values from all the Simulations Done

As we spent a significant amount of time and computation resources history matching this well, it is nature we want to seek additional value from the simulations we did, especially from the scenarios that are not HM solutions.

One of the frequently asked questions is the effect of a better or cheaper completion design on the NPV and EUR. More simulations can be performed to answer these questions, but a less reliable but cheaper result can be provided based on simulations we have already run. During the history matching, the simulation results are stored, and the production data is only used to compare with the real production history and get an evaluation on how well it matches the reality. However, that production profile can also be used to calculate the NPV for evaluation. Although many of the scenarios we designed may not be the history matching solution, some of the discrepancies may come from the completion; if that is the case, we may have an insight into the effect of the completion.

The uncertain properties for the reservoir model can be divided into two groups; one is related to the completion and the other to the first property of the reservoir. The completion-related property may include fracture half-length, fracture conductivity, fracture height, and fracture water saturation. The reservoir related property may include

matrix permeability, porosity, and water saturation. We may want to constrain the reservoir-related uncertain properties and test different completions.

As usual, the challenges come from the dimension of the uncertainty space. The first challenge is the insufficient data coverage because of the way we sampled the uncertain space. Very fewer scenarios are evaluated by the simulator. With the data available, it is hard to exclude the influence caused by the difference in reservoir properties. What makes it worse is that because the primary objective of history matching is not to build a good overall proxy model, we may not have very much points that are far from history matching solution to have a good estimation there. The second challenge is that the visualization of a high dimensional model is not easy, especially when you want to study the influence of several properties.

### 7.2 Interactive Parallel Coordinates Plots in Our Platform

The interactive parallel coordinates plot (IPCP), with the help of proxy models, are used in this study to explore the problem. The basic idea of the IPCP plot is similar to the static one that is introduced in Chapter 3.X.X, but the range of properties can be constrained to check fewer results. Figure 6.X is the IPCP of all the simulated scenarios. RMSE, NPV and 30 years productions are added to provide more information.

The IPCPs to be shown in this thesis are implemented with the Plotly library on Python. The plots are shown in HTML and can be displayed by any browser that supports JavaScript. Part of the components in the plots are not following the common industrial practice and thus are introduced here. For each coordinate, the name and unit of the parameter are listed on the top. The brief names are used here for better alignment in visualization. The label “Sum” represents the weighted RMSE. The color of the lines can be assigned colors base on one property and in the plot above, the lines are colored base on the RMSE. The lines that have one or more properties out of the range will be in light gay. The values of the parameters are also automatically converted by the plotting library. The values of multipliers behind the numbers are listed in Table 7.1.

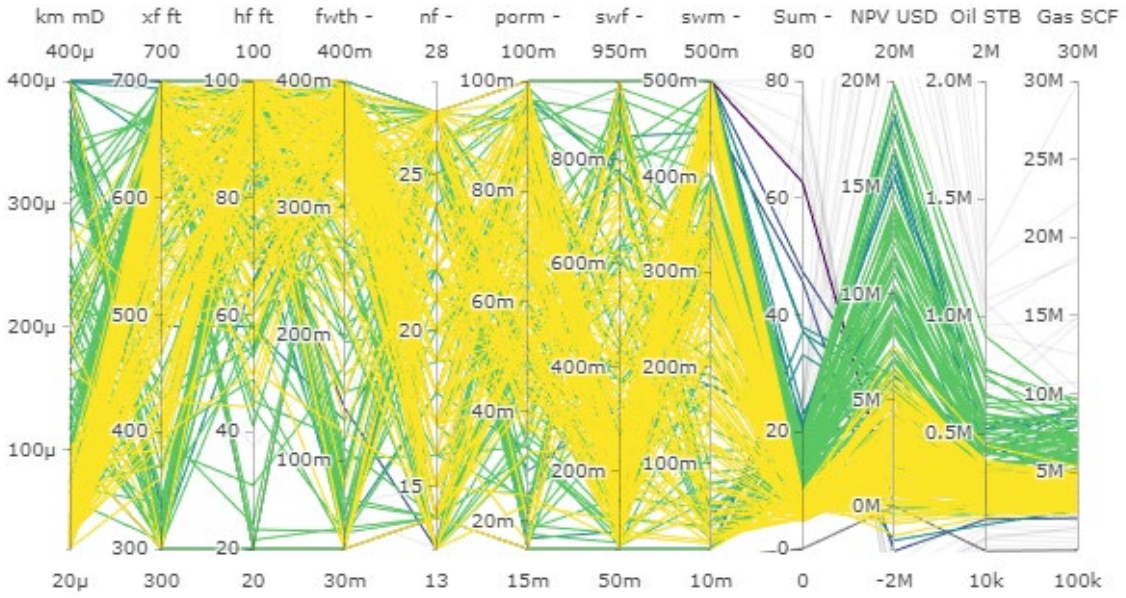



Figure 7.1: IPCP of the simulated scenarios.

Table 7.1: The value of the multiplier in the plotting area of IPCP

| **Symbol** | $\boldsymbol{\mu}$ | $\boldsymbol{m}$ | $\boldsymbol{k}$ | $\boldsymbol{M}$ |
|---|---|---|---|---|
| **Value** | $10^{-6}$ | $10^{-3}$ | $10^{3}$ | $10^{6}$ |

An example of constraining the range of a specific property is shown in Figure 7.2. After constraining the weighted RMSE to 10, only the scenarios that are regarded history matching solutions are displayed. Multiple constraints on different properties can be added, modified, and canceled to dive into the details of the dataset.

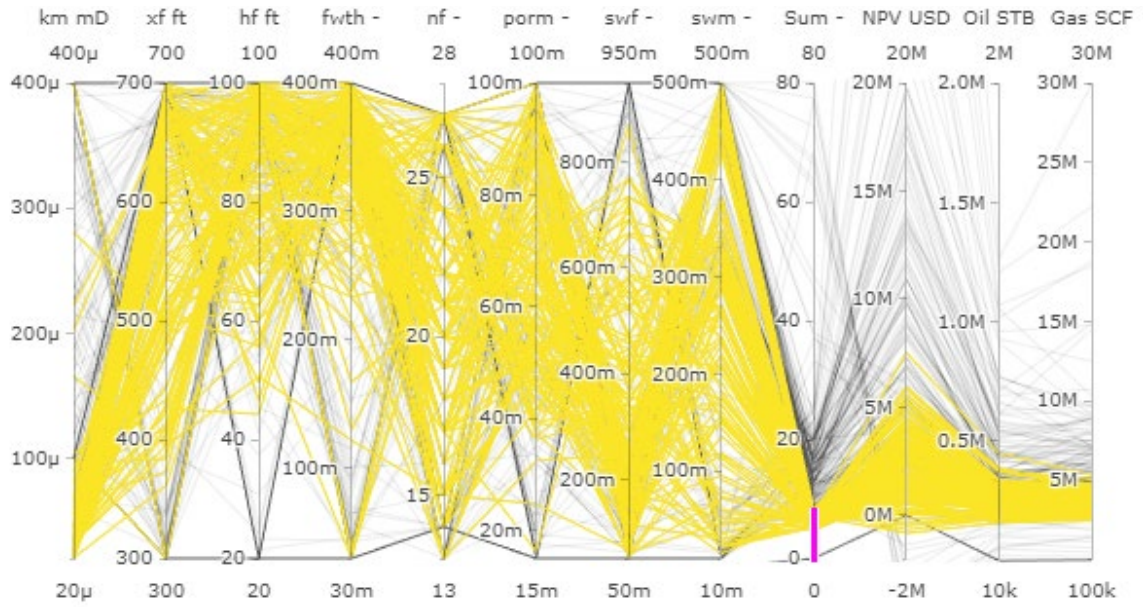
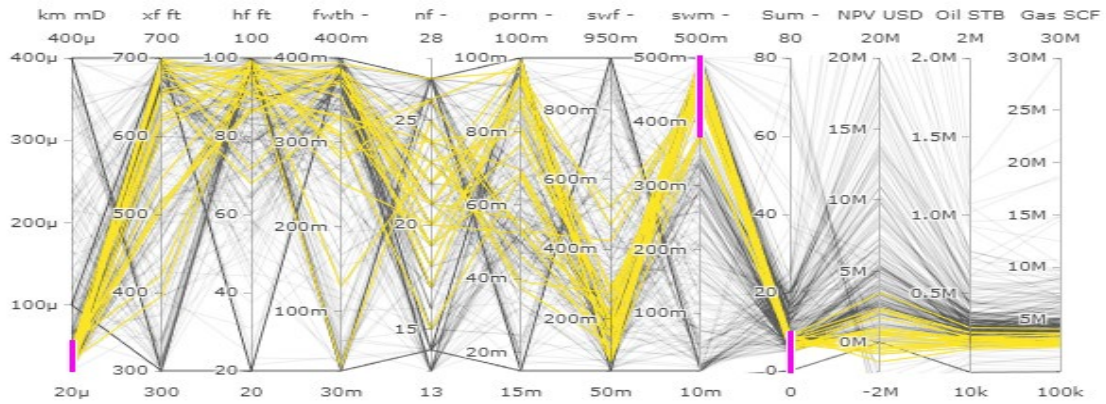



Figure 7.2: IPCP of the simulated scenarios when the RMSE is limited to the range of solution.

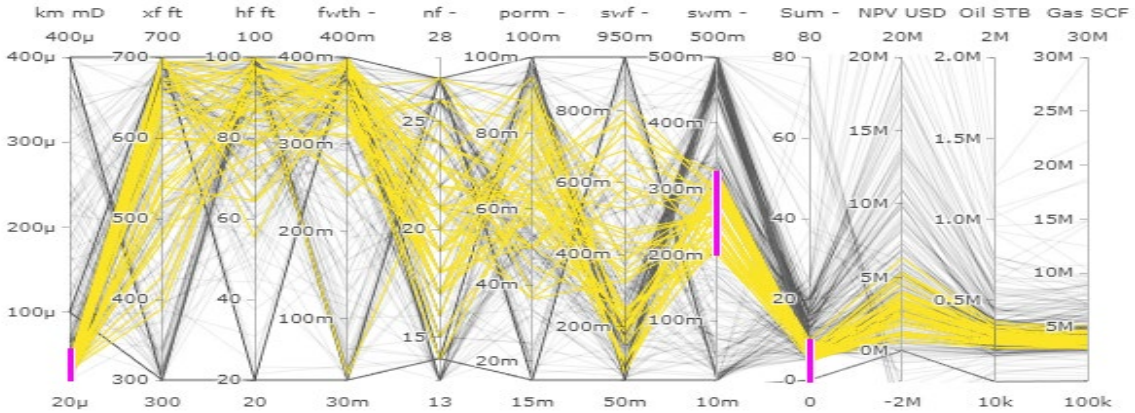
### 7.3 Directly Analyze all the Simulated Scenarios

The IPCP can be used directly to analyze the solutions of history matching. Especially when we want to understand the influence of a small number of uncertain properties.

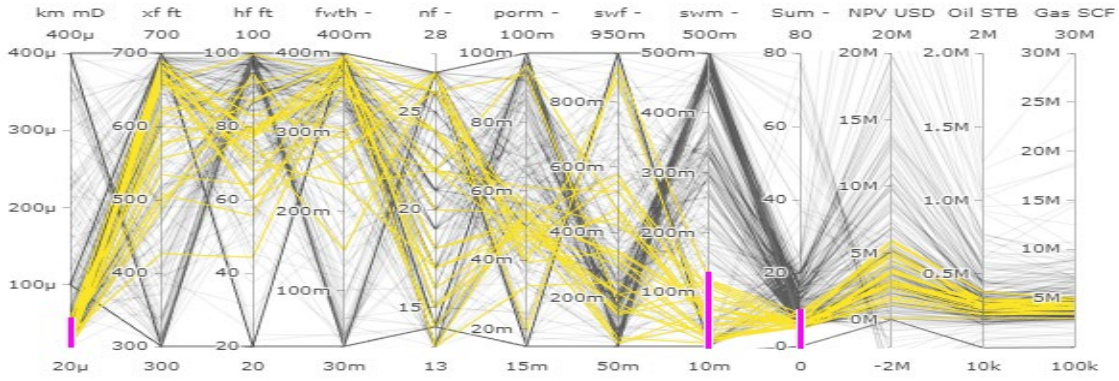
The effect of matrix water saturation is shown in Figure 7.3. The RMSE is constrained to the range of solution; the matrix permeability is also constrained to lower end. When we constrain the matrix to the upper end, in order to match the water production, the fracture water saturation is more likely to be in the lower end and the matrix porosity upper end. The results for the long-term production is that the oil and gas production is considerably lower than two other situations and has a lower NPV at the same time. Similar analyzation can be done when matrix water saturation is at the lower level or middle level. If we managed to understand the matrix water saturation from other sources, then we can directly constrain it in the IPCP for a quick assessment.




(a) High matrix water saturation

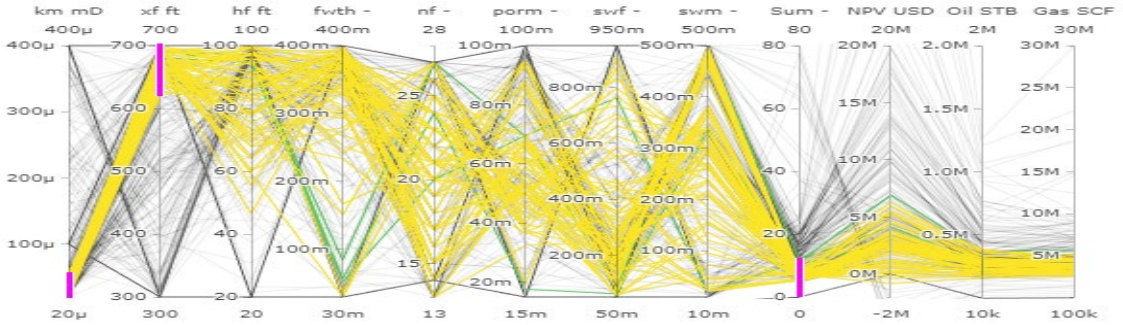



(b) Middle water saturation

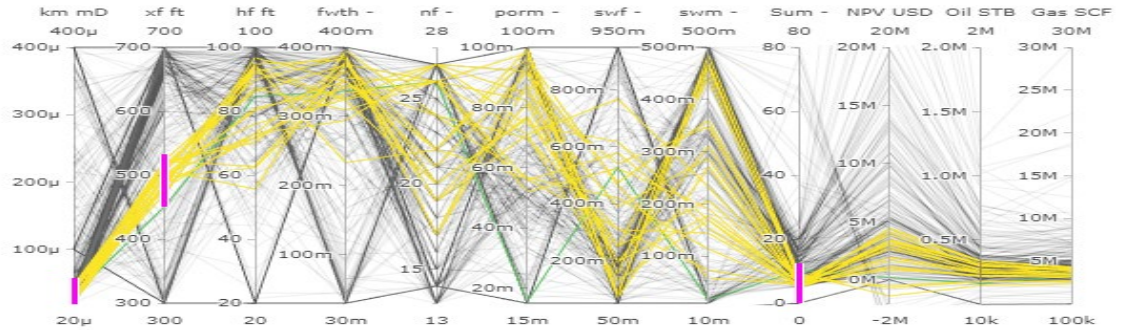



(c) Low water saturation

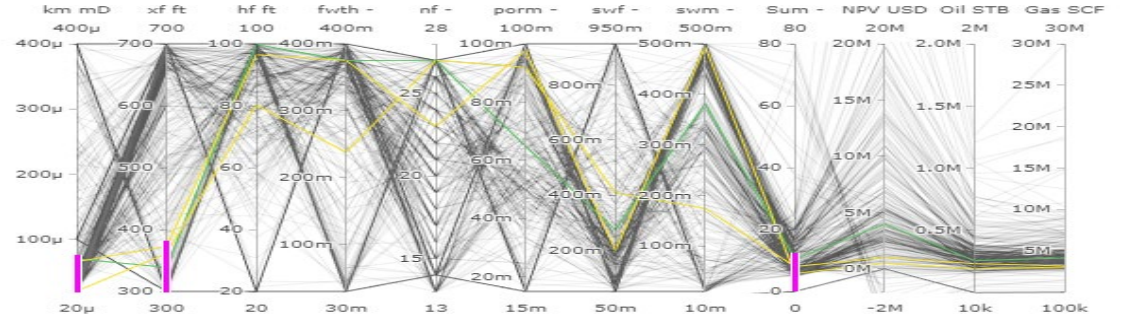
Figure 7.3: Distribution of the solution when matrix water saturation is different.




(a) Long fracture half-length




(b) Intermediate fracture half-length




(c) Short fracture half-length

Figure 7.4: Distribution of the solution when fracture half-length is different.

A similar exploration can be done to the completion related property, and an example on fracture half-length is demonstrated in Figure 7.4. For this case, we do not have much-matched scenarios when the fracture half-length is at a middle and lower level, due

to the way we designed the scenarios during the history matching process. Similar problems occurs when more constraints are added. Figure 7.5 shows an example when 4 constraints are added simultaneously. This lack of data problem will severely undermine the ability to perform a more exploratory application.

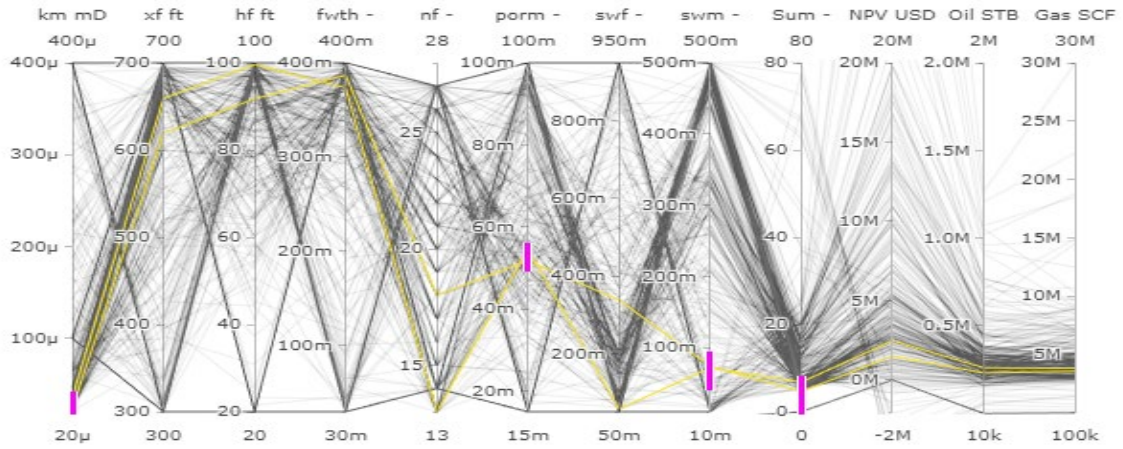



Figure 7.5: More constraints added to the IPCP.

### 7.4 Prediction of Proxy Model on Batch Scenarios

We can track down to some typical solutions by constraining the range of some uncertain properties. Then, we can cancel constrain on RMSE and check the scenarios that are similar to the solutions. However, those non-solution scenarios are very less and are distributed in an uncertain space. It is very likely not much similar scenario can be used directly. Despite the challenges we discussed in the previous chapter, the proxy model needs to be used to generate more scenarios. The basic workflow to prepare the tools is listed here.

1. Train the proxy model to predict the NPV, RMSE and the production or other simulation results of interests.
2. Design the scenarios and evaluate them with the proxy model.
3. Make an IPCP for all the properties and responses of the newly designed scenarios and another for that of original history matching cases.

### 7.4.1 Build the proxy model with data available

The response of the model will include the RMSE, NPV, and other simulation results of interests like the EUR. The predictions can be made simultaneously by ANN, and the time needed for ANN to make the batch prediction after it is properly trained is relatively low. Bagged ANN is used here to make the prediction. Consider we have a considerable small dataset and most of the scenarios are clustered around the solution, it may not be the best idea to waste the data available on developing and testing dataset. Before training the bagged ANN with the entire dataset, we need to make sure that there is no severe overfitting issue if we spare part of the data for cross-validation.

The model is trained by iteration, and the calculated R square of training, developing and testing dataset are plotted in Figure 6.6. The proxy performs well on all the dataset and thus indicating ineligible overfitting. We can also reach this conclusion by directly comparing the prediction with simulation results as shown in Figure 7.7.

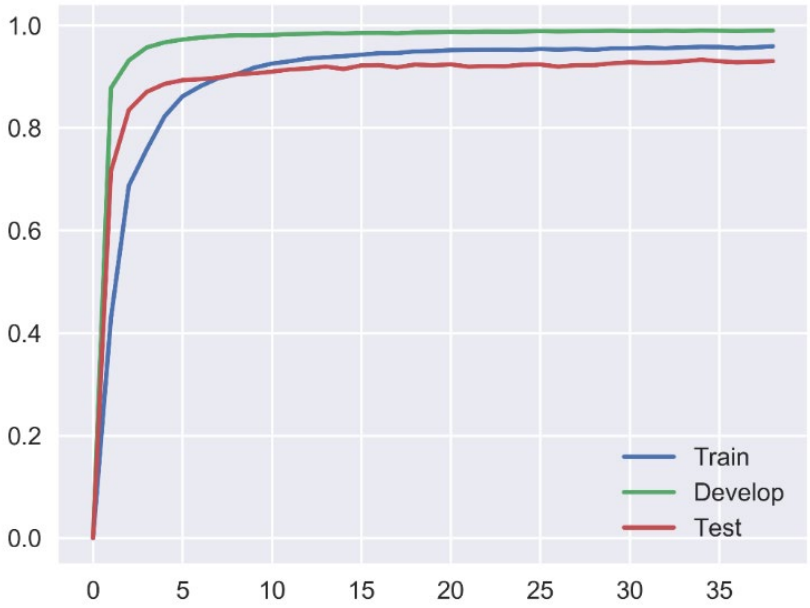
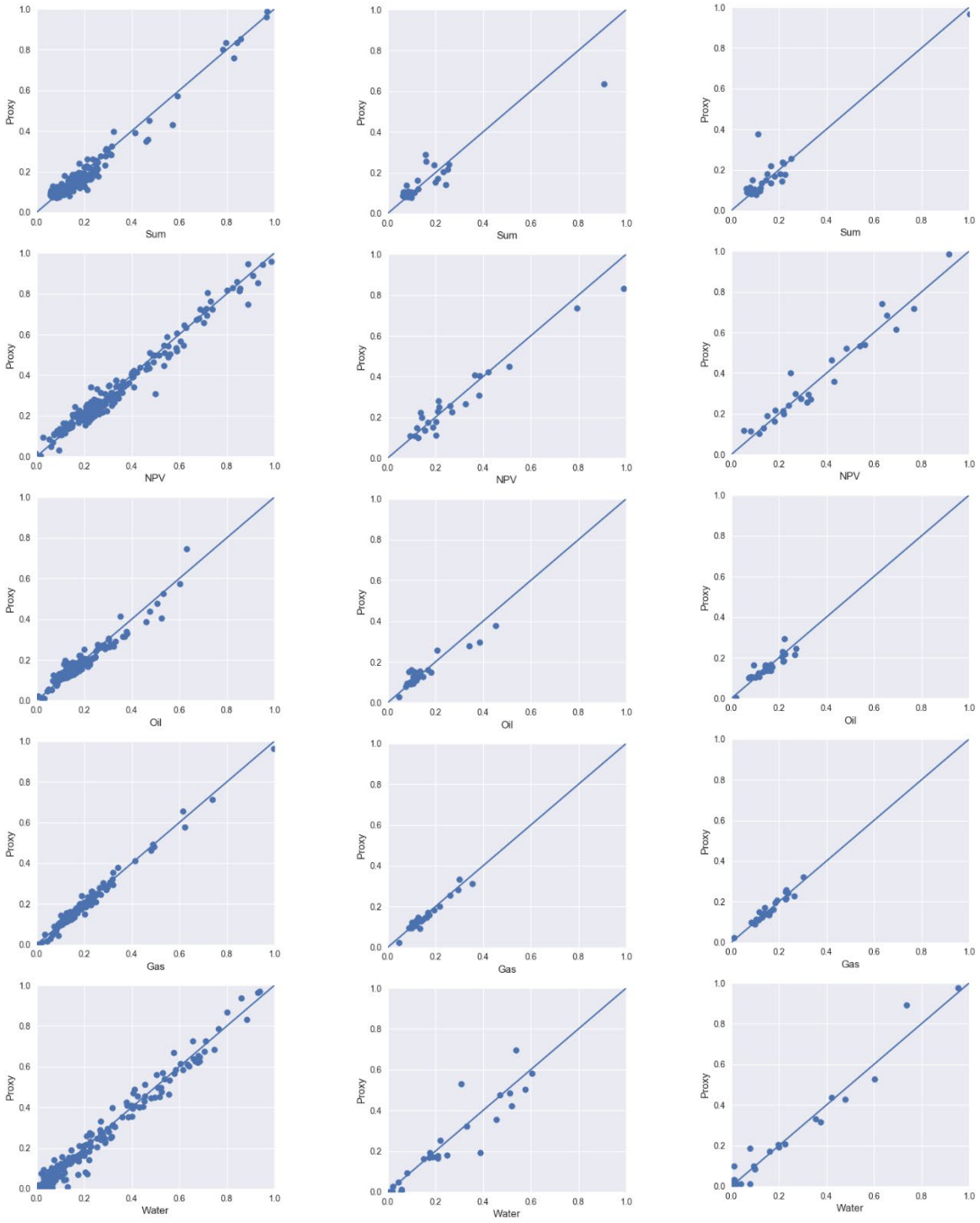



Figure 7.6: ANN's R square after different numbers of training iteration.




(a) Training dataset (b) Developing dataset (c)Testing dataset

Figure 7.7: ANN's prediction on (a) training, (b) developing and (c) testing dataset after 20 iterations of training. Make sure that the figure(s) and the title stay in the same page.

### 7.4.2 Large batch of random scenarios

A large batch of random scenarios can be used for exploratory analysis when there is no specific goal. In this section 500,000 random scenarios are evaluated by the proxy model and fed to the IPCP. Figure 7.8 shows the scenarios that are potential history matching solutions. The distribution is close to that of the scenarios evaluated by the simulator which is shown in Figure 7.2. It covers more region of the uncertain space and has a wider coverage of the response space. It is because more points can be regarded as the solutions by proxy model yet we could not run all of them during the simulation. Certainly, there would be some limitation for the model we created; we will assume ineligible errors in the study to be continued.

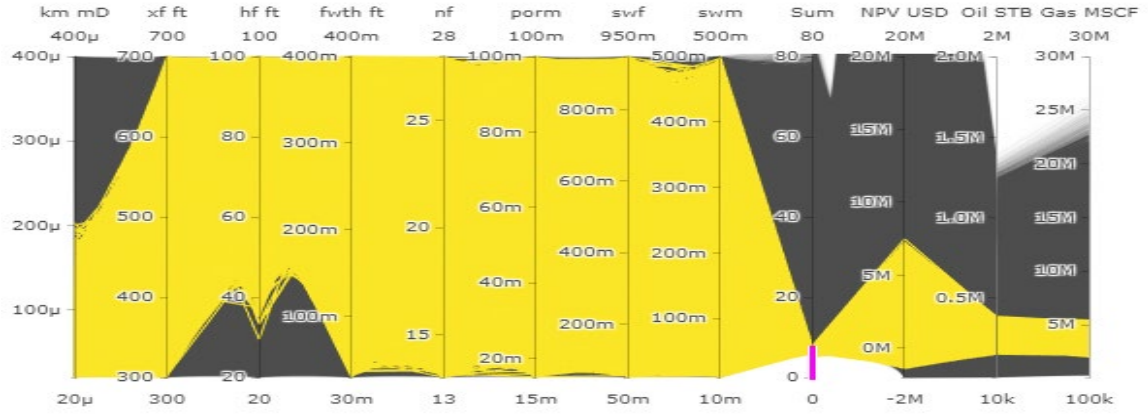
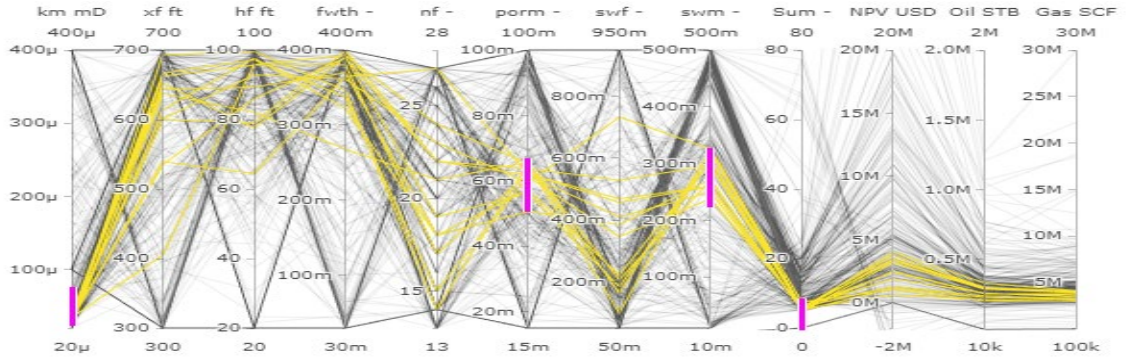
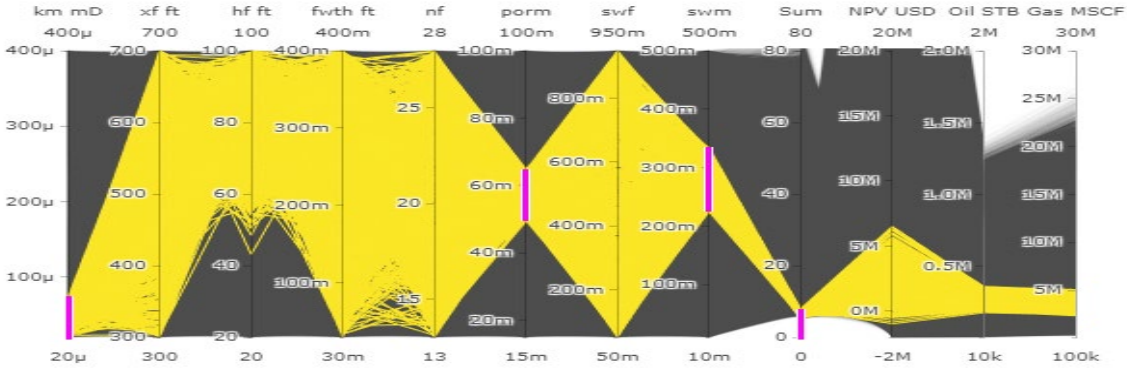



Figure 7.8: Interactive parallel coordinates plot of the proxy predicted scenarios when the RMSE is limited to the range of solution.

As can be seen in Figure 7.9, even if more constraints are added, there are still plenty of scenarios available for evaluation. In this example, the reservoir-related properties are constrained to a certain range and represent a type of realization of the reservoir. The corresponding distribution of proxy predicted NPV and EUR could be directly checked from the IPCP. It is easy to check the NPV of other reservoir realizations.



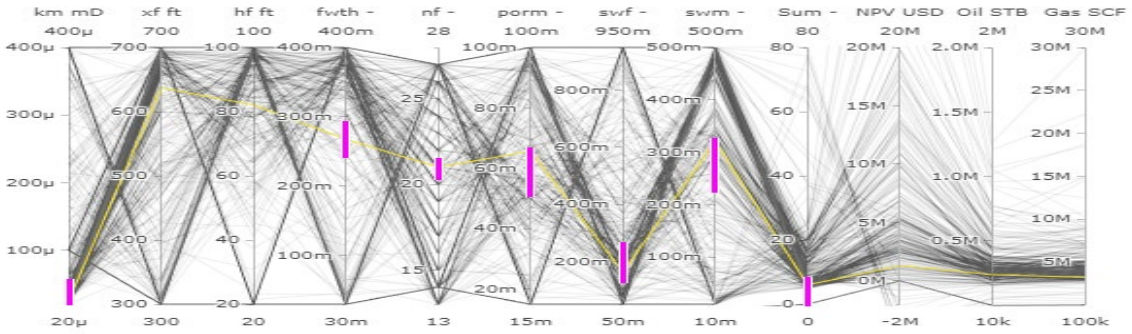

(a) Simulated scenarios



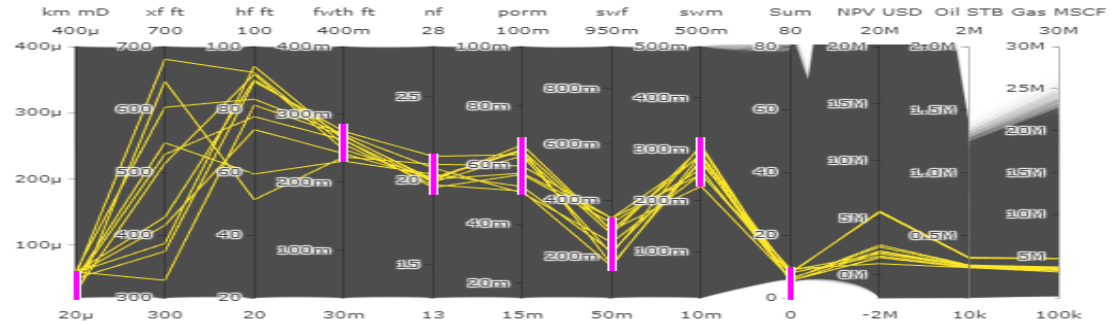

(b) Proxy predicted scenarios

Figure 7.9: Interactive parallel coordinates plot of the simulated scenarios and proxy predicted scenarios when the RMSE is limited to a specific range.

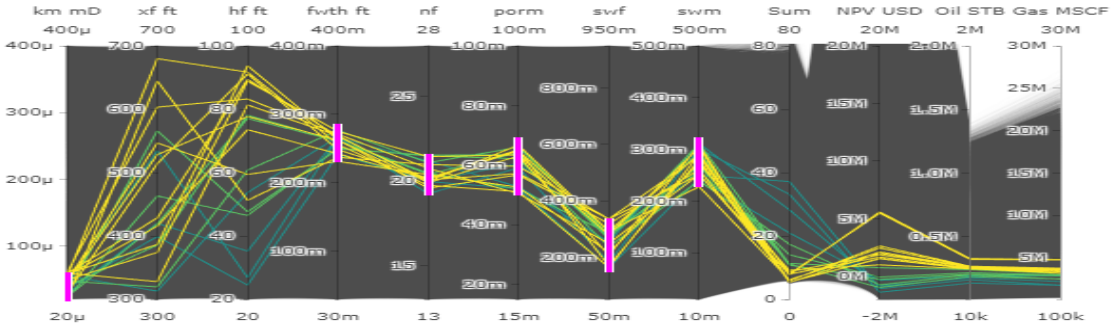
The IPCP of proxy evaluated scenarios can be used together with the simulator evaluated scenarios to explore the effect of one or two parameters on a specific realization. As shown in the Figure 7.10, if we look into one scenario that is regarded the history matching solution, the number of scenarios still cannot provide good coverage even if the constraint on RMSE is canceled. We can still trace the lines in the last plot and get the conclusion that fracture height has a larger effect than the fracture half-length. It is still difficult to draw a safe conclusion since the range of constraints are wide and can be the cause of the differences.




(a) Simulated scenarios when RMSE is not constrained




(b) Proxy predicted scenarios when RMSE is constrained

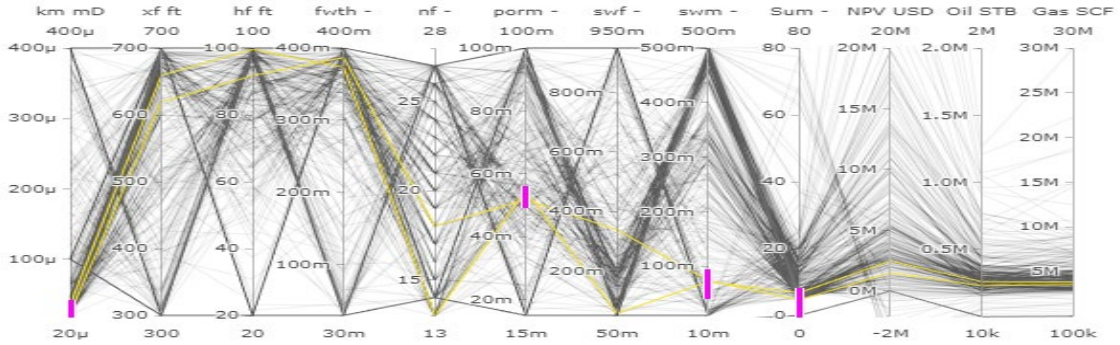



(c) Proxy predicted scenarios without constraints on RMSE

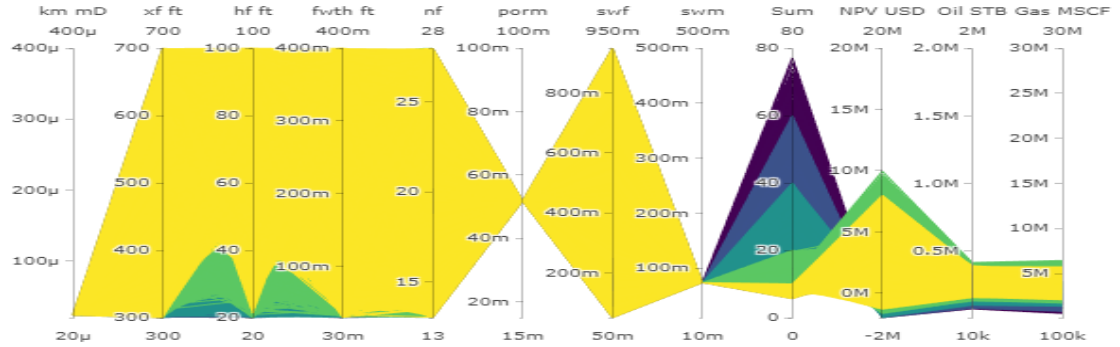
Figure 7.10: Interactive parallel coordinates plot of the simulated scenarios and proxy predicted scenarios when several properties are limited to a specific range.

### 7.4.3 Proxy prediction on a specific reservoir realization

To exclude the effects caused by the controlled properties, the reservoir-related properties are set to the exact value of a specific reservoir realization. Other properties are randomly generated. The exact solution to be shown in this example is visualized in Figure 7.11.




(a) Selected simulated scenarios that matched the production history




(b) Proxy predicted scenarios without constraints on RMSE

Figure 7.11: Designed scenarios that have the same realization of reservoir properties.

Since the economy is of priority interest in this section, the plots in the following part are adjusted accordingly. The color of the lines represents a different level of NPV, and the corresponding part can be referred to the coordinate of NPV in the plot. The plotting range of each response is adjusted according to the range of predictions. The results of all the scenarios are shown in Figure 7.12.

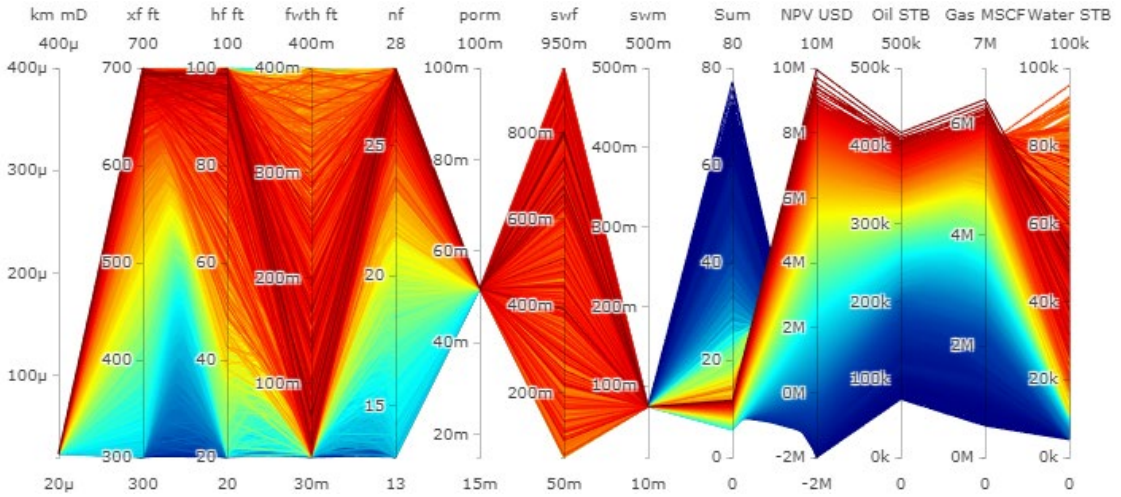
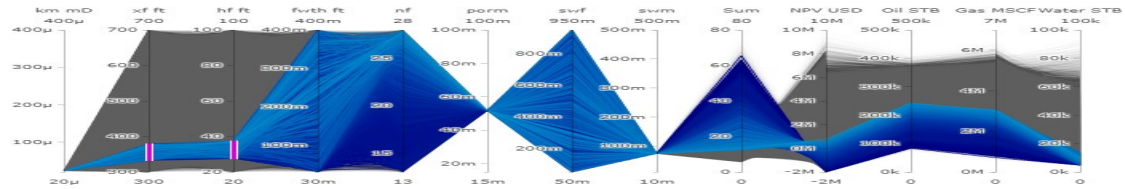
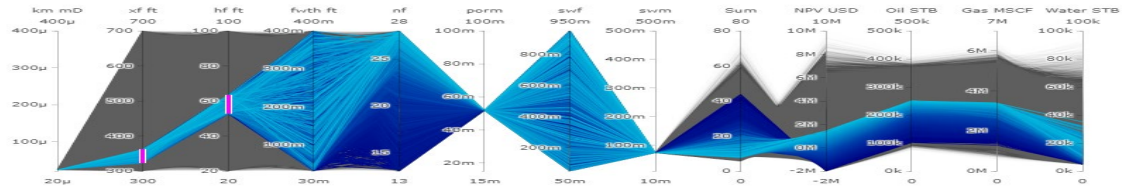


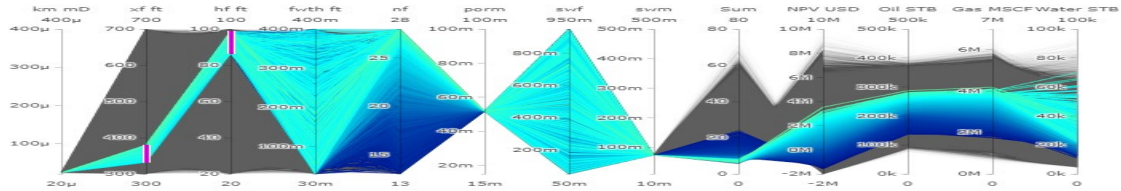

Figure 7.12: Designed scenarios that have the same realization of reservoir properties.



(a) Low fracture half-length and low fracture height

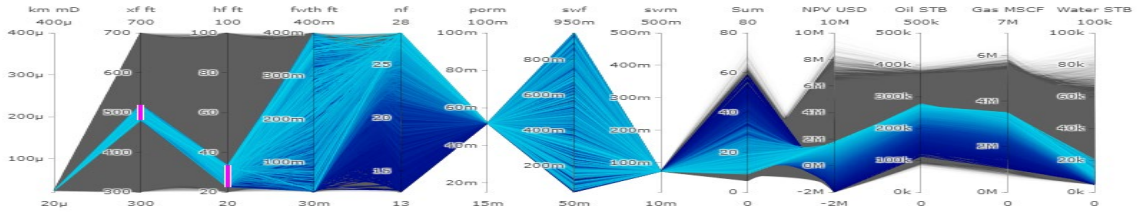


(b) Low fracture half-length and middle fracture height

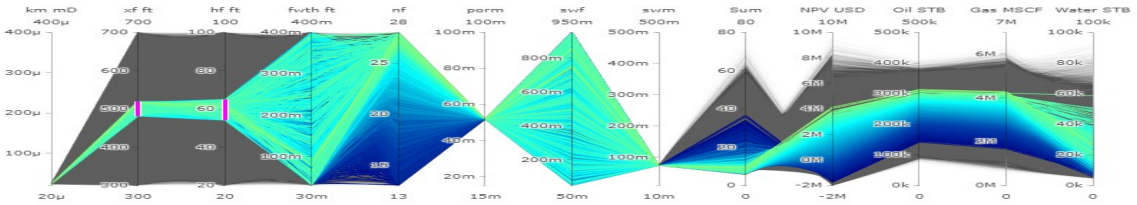


(c) Low fracture half-length and high fracture height

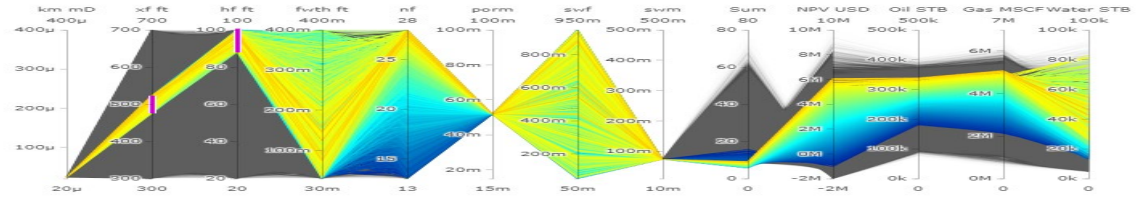
Figure 7.13: Effect of fracture half-length and fracture height on NPV and EUR.



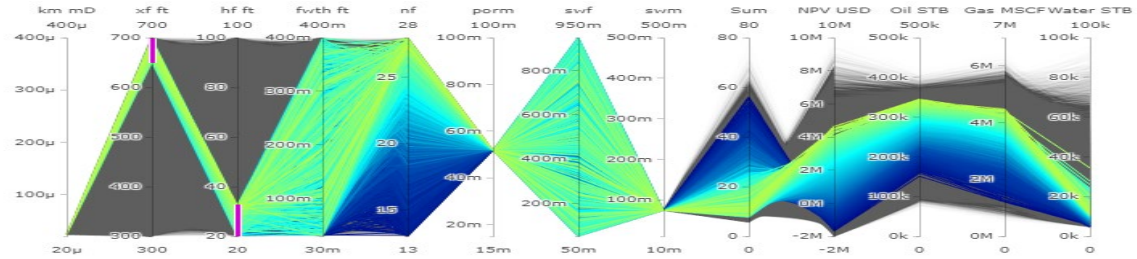

(d) Middle fracture half-length and low fracture height



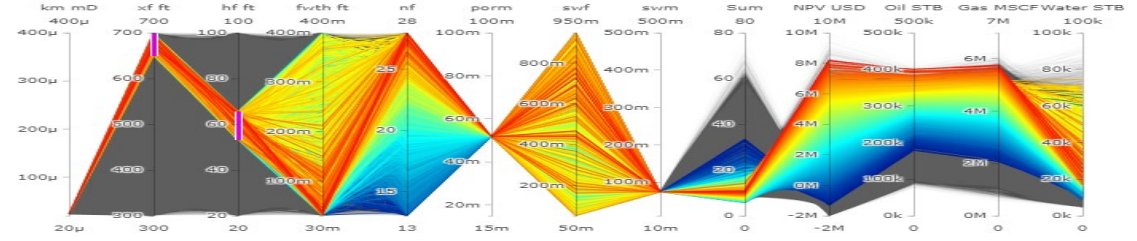

(e) Middle fracture half-length and low fracture height




(f) Middle fracture half-length and high fracture height

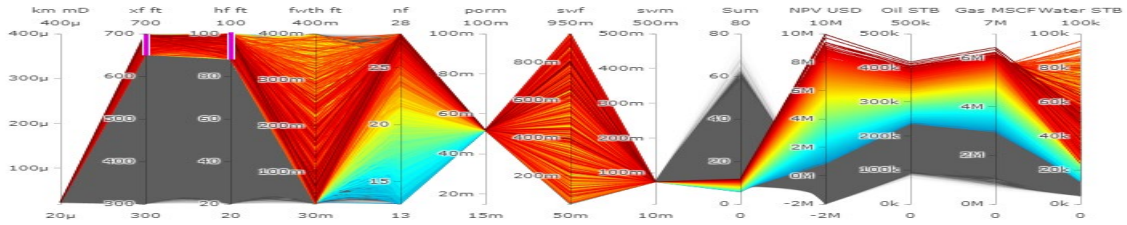



(g) High fracture half-length and low fracture height




(h) High fracture half-length and middle fracture height

Figure 7.13 continued




(i) High fracture half-length and high fracture height

Figure 7.13 continued

### 7.4.4 Proxy prediction on representative reservoir realization

The primary issue of previous analysis is that the chosen realization of the reservoir may not be representative enough. Generating the possible realization one after another will require time to distinguish and understand them. Thus, the concept of a representative scenario is used again. For this case, the simulated scenarios are clustered into 10 groups based on the normalized reservoir related uncertain properties, and the scenario with the lowest RMSE is chosen as the representative solution. The detailed distribution of each reservoir is shown in Figure 7.13. For each reservoir realization, 200,000 scenarios are generated and evaluated by the proxy model. All the scenarios are presented in Figure 7.14 for economic analysis.

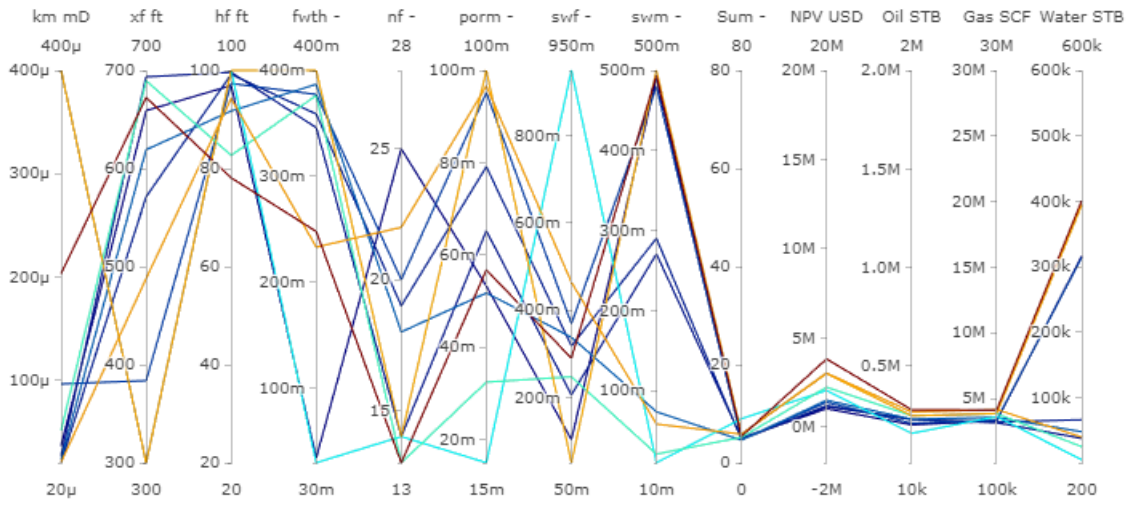



Figure 7.14: Representative history matching solution based on clustering.

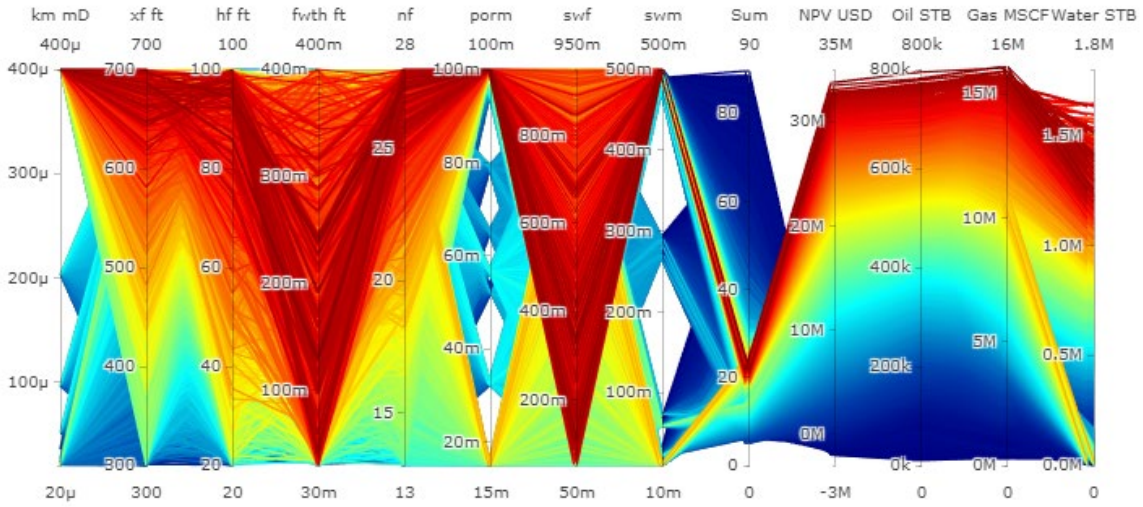



Figure 7.15: Designed scenarios that have the same realization of reservoir properties.

Similar to the previous study, we can limit the range of properties and responses for a different purpose. Figure 7.15 is an example to check the economic performance of potential solutions for chosen representative reservoir realizations. Figure 7.16 is an example that estimates the fracture half-length's effect on the economy of multiple reservoir realizations.

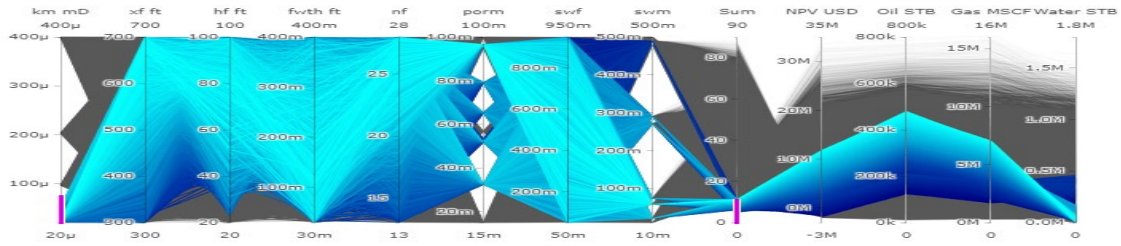
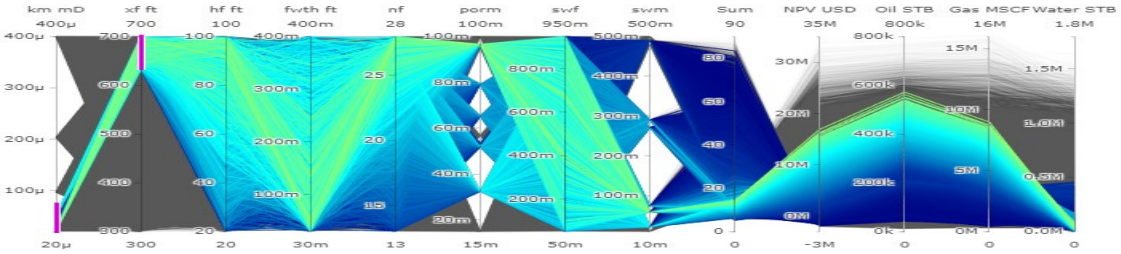
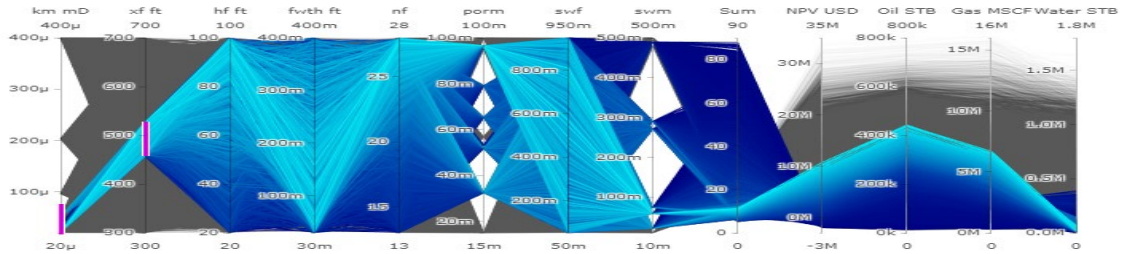


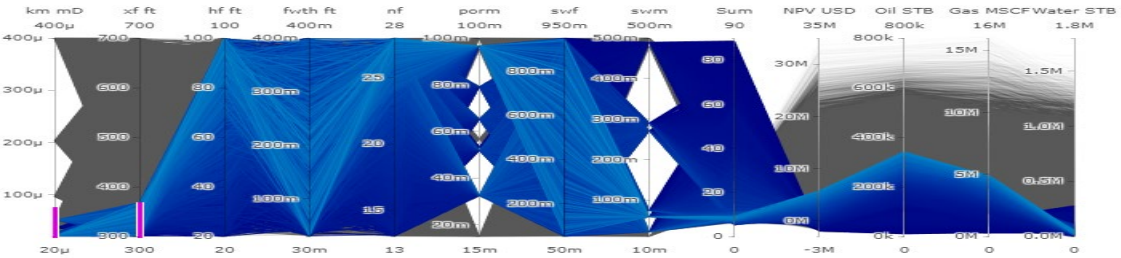

Figure 7.16: Potential solutions for chosen representative reservoir realizations.




(a) High fracture half-length



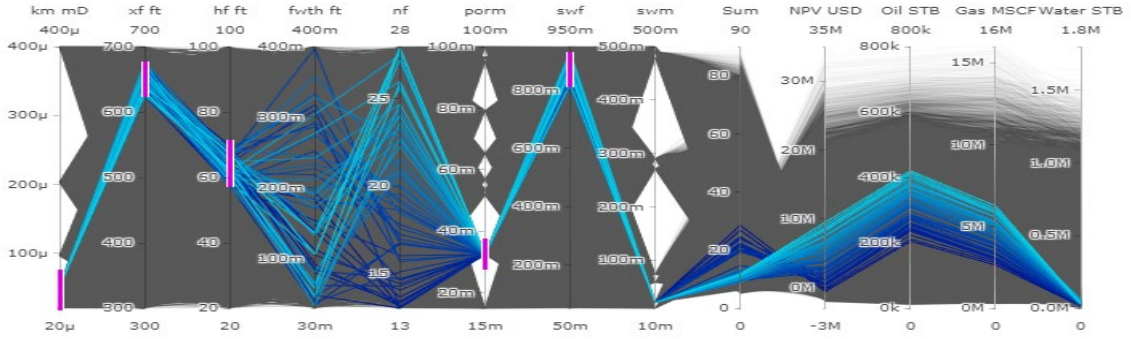

(b) Middle fracture half-length



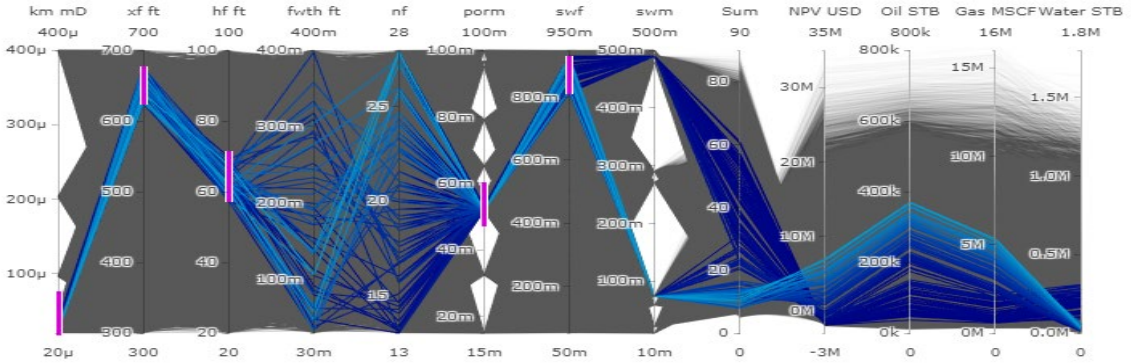

(c) Low fracture half-length

Figure 7.17: Effect of fracture half-length for chosen representative reservoir realizations.

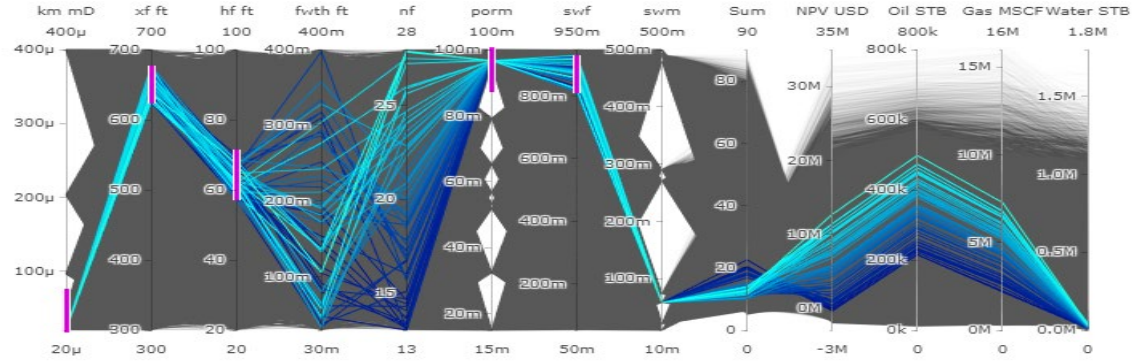
It is also easy to go to the detail of a specific solution by constraining the reservoir property to a narrow range. As shown in Figure 7.17, the coverage of unspecified property is good enough to work on.




(a) Low porosity




(b) Middle porosity




(c) High porosity

Figure 7.18: Investigate a specific reservoir realization.

## 7.5 POTENTIALS

For this project, we have 9 uncertain property, which makes it hard to cover the entire uncertainty space. Taking the memory and the ability of browser into account, 500,000 scenarios are generated and evaluated. Moreover, since so many cases are to be evaluated, we need the proxy model that cost reasonable time. ANN and KNN are both capable of predicting multiple responses at the same time, and the computation time is low due to the vectorization. ANN can capture nonlinear behavior and thus is the preferred model here. KNN, due to its calculation mechanic, is more for interpolation and its prediction is relatively conservative.

The process of training the ANN for this problem is hard with the 272 cases at hand; we want to take full advantages of them rather than spare some for the developing and testing set. However, it is not a bad idea for us to follow the common practice while iterating for preferable training times and ANN structure. To find an ANN structure that is preferred for this problem, the workflow in chapter 4 is used. Then this structure is bagged in the continuous training of our model.

We also found that the weighted RMSE is harder to predict than other property, yet using a single model to predict is not making a significant improvement. The major reason is that our sampling is more concentrate on the low RMSE region. This is made worse since the nonlinearity of RMSE is way higher than the productions. If this kind of method is to be explored in the future, we definitely need to reconsider how we should design the scenarios in the history matching phase to build a better proxy model.

One thing we always need to keep in mind is that the method used in this study is more like intuitive tools that provide a helpful understanding for people who know nearly nothing, more reliable results will require more simulation on well-designed scenarios.

# Chapter 8: Summary, Conclusion, and Recommendations

In this chapter, the summary and conclusions for this thesis are presented, recommendations for future works are provided.

## 8.1 Using Proxy Model in the Workflow

Using the proxy models in the workflows can help save the resources spent on simulation, yet different levels of accuracy are required for different problems. Although a proxy with high accuracy is always welcomed, we could still benefit from the proxy model that is less accurate but have less cost.

Using the model while exploring the possible scenarios for history matching does not require a proxy model that predicts the exact value, on the country, it will care more about the relative size to locate the potential better scenario. Also, as a part of the automated iterative process, spending too many resources training the model or improve and select the model may not worth the efforts.

Higher accuracy may be helpful when the proxy model is used for the Metropolis-Hasting process. Although it is still the relative size, we care more about, the fact that we only need to build the model once makes tuning the model more affordable.

Considerable accuracy is necessary when we use the proxy model to predict the NPV. For this case, the model is used to predict the exact value. Since economic decisions are to be made base on the results, it worth every penny tuning this model.

When the proxy model is used in another situation, it is also a good habit to decide how much resources to allocate based on the purpose it is to serve.

## 8.2 Scenario Design

In this thesis, the uncertainties are studied base on a large number of possible scenarios. How to generate the scenarios and how to evaluate the scenarios are two

significant problems we are facing. To solve the ultimate problem, scenarios are designed in the workflows for different purposes. In the platform, the scenario is designed by the sampling module and the sampling methods for different objectives are discussed in section 3.2. The goal of the entire project and the goals at each step shall be balanced during the process. If a proxy model is to be built for further analysis, the diversity of scenarios shall be considered in previous steps. For the sampling process with multiple objectives, the sampling units, and sampling strategies are promoted. Similar to the example in section 6.4.1, governing equations can be used to control the strategy by deciding the input, and output amount of sampling units.

## 8.3 One Step Beyond the History Matching

Digging as much value as possible from the history matching results is important, as it may take lots of resources. The examples in Chapter 6 provides a workflow to optimize the well spacing plan by simulating the selected history-matched scenarios with different well spacing models. The platform enables the automation of new multiwell models with modest adjustment to the history matching model. The boxplots showed the trend of NPV when the spacing plan is adjusted for the reservoir under uncertainty. The example in Chapter 6 estimated the posterior distribution of uncertain properties with MCMC and showed a way to select representative history matching solutions for further simulation. Then all the scenarios evaluated by simulators are fed to the interactive parallel coordinate plot in Chapter 7. The NPV and EUR are shown together in IPCP for decision-maker to look into the scenarios of interest. With the help of a well-trained proxy model, an unlimited number of scenarios can be evaluated and play with for different studies.

## 8.4 Platform Usage

The platform is designed to automate the workflows that help evaluate the unconventional assets. The calculation interface of the platform provides an easy and universal way to generate complex simulator input files by batch and make the automation easier. The implemented ANN in the module of statistical models also provides an easy

way to train and select the model to perform statistical learning. Two examples of history matching and one example of well spacing are demonstrated in the thesis. More workflows that are potentially useful are discussed.

## 8.5 RECOMMENDATIONS FOR FUTURE WORK

In this study, potentially useful tools to manage the unconventional reservoir are explored from an engineering point of view. Most of the implementations and ideas are based on intuition and experience rather than strict derivation. Moreover, the potential of those ideas has not been fully explored. Here some of the recommendations for future work are listed.

Include the geomechanic part in the simulation to evaluate the possible influence of nearby wells.

Improve the efficiency of training and using ANN to reduce the time needed.

Develop a theory to guide the governing equation of sampling strategies.

A better theory of normalizing the uncertain variables like the permeability and fracture width.

# References

Alfarge D, Wei M, Bai B. 2018. Integrated Investigation of CO 2-EOR Mechanisms in Huff-n-Puff Operations Based on History Matching Results. Paper SPE 190234, presented at SPE Improved Oil Recovery Conference, Tulsa, Oklahoma, 14-18 April.

Baca, R.G., Arnett R.C., and Langford D.W. 1984. Modelling Fluid Flow in Fractured-Porous Rock Masses by Finite-Element Techniques. International Journal for Numerical Methods in Fluids 4 (4): 337–348.

Blaskovich, F.T., Cain, G.M., Sonier, F. et al. 1983. A Multicomponent Isothermal System for Efficient Reservoir Simulation. Paper SPE 11480, presented at the Middle East Oil Technical Conference and Exhibition, Manama, Bahrain, 14-17 March.

Chen, C., Gao, G., Li, R., Cao, R., Chen, T., Vink, J. C., and Gelderblom, P. 2017. Integration of Distributed Gauss-Newton with Randomized Maximum Likelihood Method for Uncertainty Quantification of Reservoir Performance. Paper SPE 182639, presented at SPE Reservoir Simulation Conference, Montgomery, Texas, 20-22 February.

Dachanuwattana, S., Xia, Z., Yu, W., Qu, L., Wang, P., Liu, W., Miao, J., and Sepehrnoori, K. 2018. Application of Proxy-based MCMC and EDFM to History Match a Shale Gas Condensate Well. Journal of Petroleum Science and Engineering 167: 486-497.

Dos Santos Sousa, E.P. and Reynolds, A.C. 2019. Markov Chain Monte Carlo Uncertainty Quantification with a Least-Squares Support Vector Regression Proxy. Paper SPE 193918, presented at SPE Reservoir Simulation Conference, Galveston, Texas, 10-11 April.

Dramsch, J.S. and Lüthje, M., 2018. Deep-learning seismic facies on state-of-the-art CNN architectures. Paper SEG 2018-2996783, presented at SEG International Exposition and Annual Meeting, Anaheim, California, 14-19 October.

Emerick, A. A., and Reynolds, A. C. 2012. Combining the Ensemble Kalman Filter With Markov-Chain Monte Carlo for Improved History Matching and Uncertainty Characterization. Paper SPE 141336, presented at SPE Reservoir Simulation Symposium, The Woodlands, Texas, 21-23 February.

Eltahan, E., Yu, W., Sepehrnoori, K., Kerr, E., Miao, J., and Ambrose, R. 2019. Modeling Naturally and Hydraulically Fractured Reservoirs with Artificial Intelligence and Assisted History Matching Methods Using Physics-Based Simulators. Paper SPE 195269, presented at SPE Western Regional Meeting, San Jose, California, 23-26 April.

Erbas, D., and Christie, M. A. 2007. Effect of Sampling Strategies on Prediction Uncertainty Estimation. Paper SPE 106229, presented at SPE Reservoir Simulation Symposium, Houston, Texas, 26-28 February.

Ertekin T., Abou-Kassem J. H., King G. R., 2001. Basic Applied Reservoir Simulation. SPE Textbook Series Vol. 7, Society of Petroleum Engineers.

Fiallos Torres, M.X., Yu, W., Ganjdanesh, R., Kerr, E., Sepehrnoori, K., Miao, J., Ambrose, R. 2019a. Modeling Interwell Fracture Interference and Huff-N-Puff Pressure Containment in Eagle Ford Using EDFM. Paper SPE 195240, presented at SPE Oklahoma City Oil and Gas Symposium, Oklahoma City, Oklahoma, 9-10 April.

Fiallos Torres, M.X., Yu, W., Ganjdanesh, R., Kerr, E., Sepehrnoori, K., Miao, J., Ambrose, R. 2019b. Modeling Interwell Interference Due to Complex Fracture Hits in Eagle Ford Using EDFM. Paper IPTC 19468, presented at International Petroleum Technology Conference, Beijing, China, 26-28 March.

Forsyth, D., Al Musharfi, N. M., Al Marzooq, A. M., 2011. Tight Gas Petrophysical Challenges in Saudi Aramco. Paper SPE 149048-MS, presented at SPE/DGS Saudi Arabia Section Technical Symposium and Exhibition, Al-Khobar, Saudi Arabia, 15-18 May.

Frash, L. P., Hood, J., Gutierrez, M., Huang, H., Mattson, E., 2014. Laboratory Measurement of Critical State Hydraulic Fracture Geometry. Paper ARMA-2014-7316, presented at 48th U.S. Rock Mechanics/Geomechanics Symposium, Minneapolis, Minnesota, 1-4 June.

Gang, T., and Kelkar, M.G. 2008. History Matching for Determination of Fracture Permeability and and Capillary Pressure. SPE Reservoir Evaluation and Engineering 11 (5): 813 – 822.

Gao, G., Vink, J. C., Chen, C., Tarrahi, M., and El Khamra, Y. 2016. Uncertainty Quantification for History Matching Problems with Multiple Best Matches Using a Distributed Gauss-Newton Method. Paper SPE 18161, presented at SPE Annual Technical Conference and Exhibition, Dubai, UAE, 26-28 September.

Goodwin, N. 2015 Briding Gap between Deterministic and Probabilistic Uncertainty Quantification Using Advanced Proxy Based Methods. Paper presented at SPE 173301, SPE Reservoir Simulaiton Symposium, Houston, 23-25 February.

Guo, Z., Chen, C., Gao, G., and Vink, J. 2017. Applying Support Vector Regression to Reduce the Effect of Numerical Noise and Enhance the Performance of History Matching. Paper SPE-187430, presented at SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 9-11 October.

Hanea, R., Evensen, G., Hustoft, L., Ek, T., Chitu, A., and Wilschut, F. 2015. Reservoir Management Under Geological Uncertainty Using Fast Model Update. Paper SPE 173305, SPE Reservoir Simulation Symposium, Houston, Texas, 23-25 February.

Hamdi, H., Couckuyt, I., Sousa, M.C. and Dhaene, T., 2017. Gaussian Processes for history-matching: application to an unconventional gas reservoir. Computational Geosciences, 21(2), pp.267-287.

Hegde, C., Wallace, S., and Gray, K. 2015. Using Trees, Bagging, and Random Forests to Predict Rate of Penetration During Drilling. Paper SPE-176792, presented at SPE

Middle East Intelligent Oil and Gas Conference and Exhibition, Abu Dhabi, UAE, 15-16 September.

Jennings, A. R., Westerman, C., Westerman-Tadlock, D., Westerman, R., Anderson, M., 2006. A Systematic Approach to Improved Success With Hydraulic Fracturing Applications. Paper SPE 101837-MS, presented at SPE Annual Technical Conference and Exhibition, San Antonio, Texas, September 24-27.

Karimi-Fard, M., Durlofsky, L. J. and Aziz, K. 2004. An Efficient Discrete-Fracture Model Applicable for General-Purpose Reservoir Simulators. SPE Journal 9 (2): 227-236.

Karimi-Fard, M. and Firoozabadi, A. 2003. Numerical Simulation of Water Injection in Fractured Media using the Discrete-Fractured Model and the Galerkin Method. SPE Reservoir Evaluation and Engineering 6 (2): 117-126.

Kim, J.-G., and Deo, M.D. 2000. Finite Element, Discrete-Fracture Model for Multiphase Flow in Porous Media. AIChE Journal 46 (6): 1120–1130.

Landa, J. L., and Güyagüler, B. 2003. A Methodology for History Matching and the Assessment of Uncertainties Associated with Flow Prediction. Paper SPE-84465, presented at SPE Annual Technical Conference and Exhibition, Denver, Colorado, 5-8 October.

Lemonnier, P. and Bourbiaux, B. 2010. Simulation of Naturally Fractured Reservoirs. State of the Art. Oil Gas Sci. Technol. 65 (2): 239-262.

Luo, G., Tian, Y., Sharma, A., and Ehlig-Economides, C. 2019. Eagle Ford Well Insights Using Data-Driven Approaches. Paper IPTC-19260, presented at International Petroleum Technology Conference, Beijing, China, 26-28 March.

Medavarapu, K., Mahato, P. K., Das, S., Singh, S., Patel, K. C., Nandan, A., 2012. Production Enhancement by Customised Hydraulic Fracturing Operations: Success Stories of Gamij Field. Paper SPE 154806-MS, presented at SPE Oil and Gas India Conference and Exhibition, Mumbai, India, March 28-30.

Moinfar, A., Varavei, A., Sepehrnoori, K. et al. 2014. Development of an Efficient Embedded Discrete Fracture Model for 3D Compositional Reservoir Simulation in Fractured Reservoirs. SPE Journal 19 (2): 289-303.

Monteagudo, J. and Firoozabadi, A. 2004. Control-Volume Method for NumericalSimulation of Two-Phase Immiscible Flow in Two- and Three-DimensionalDiscrete-Fractured Media. Water resources research 40 (7): 1-20.

Murtaza, M., Al Naeim, S., Waleed, A., 2013. Design and Evaluation of Hydraulic Fracturing in Tight Gas Reservoirs. Paper SPE 168100-MS, presented at SPE Saudi Arabia Section Technical Symposium and Exhibition, Al-Khobar, Saudi Arabia, May 19- 22.

Noorishad, J. and Mehran, M. 1982. An Upstream Finite Element Method for Solution of Transient Transport Equation in Fractured Porous Media. Water resources research 18 (3): 588- 596.

Nordgren, R. P. 1972. Propagation of a Vertical Hydraulic Fracture. Society of Petroleum Engineers Journal 12(04):306-14.

Oliver, D. S. and Chen, Y. 2011. Recent Progress on Reservoir History Matching: A Review. Computational Geosciences 15 (1): 185–221.

Özdogan, U., and Horne, R. N. 2004. Optimization of Well Placement with a History Matching Approach. Paper SPE-90091, presented at SPE Annual Technical Conference and Exhibition, September, Houston, 26-29.

Perkins, T. K., and Kern, L. R. 1961. Widths of Hydraulic Fractures. Journal of Petroleum Technology, 13(09), 937-949.

Ramgulam, A., Ertekin, T., & Flemings, P. B. 2006. Utilization of artificial neural networks in the optimization of history matching (Doctoral dissertation, Pennsylvania State University).

Schuetter, J., Mishra, S., Zhong, M., LaFollette, R., 2015. Data Analytics for Production Optimization in Unconventional Reservoirs. Paper SPE 178653-MS, presented at Unconventional Resources Technology Conference, San Antonio, Texas, July 20-22.

Schulze-Riegert, R., Magdeyev, I., Komin, M., and Chernyak, V. 2017. Brownfield Development Optimization under Uncertainty – A Structured Workflow Design for Complex Case Scenarios. Paper SPE-187856, presented at SPE Russian Petroleum Technology Conference, Moscow, Russia, 16-18 October.

Shahkarami, A.,Mohaghegh, S. D.,Gholami, V., andamp; Haghighat, S. A. 2014. Artificial Intelligence (AI) Assisted History Matching. Paper SPE-169507, presented at SPE Western North American and Rocky Mountain Joint Meeting, Denver, Colorado, 17-18 April.

Shakiba, M. and Sepehrnoori, K. 2015. Using Embedded Discrete Fracture Model (EDFM) and Microseismic Monitoring Data to Characterize the Complex Hydraulic Fracture Networks. Paper SPE 175142, presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas, 28-30 September.

Shaoul, J. R., Ross, M. J., Spitzer, W. J., Wheaton, S. R., Mayland, P. J., Singh, A. P., 2007. Massive Hydraulic Fracturing Unlocks Deep Tight Gas Reserves in India. Paper SPE 107337-MS, European Formation Damage Conference, Scheveningen, The Netherlands, May 30-June 1.

Siripatrachai, N.,Rana, S.,Bodipat, K., andamp; Ertekin, T. 2014. An effective coupling of type curves and expert systems for evaluating multi-stage hydraulically fractured horizontal wells in composite dual-porosity shale gas reservoirs. Paper SPE-170963, presented at SPE Annual Technical Conference and Exhibition, Amsterdam, The Netherlands, 27-29 October.

Slotte, P. A., and Smorgrav, E. 2008. Response Surface Methodology Approach for History Matching and Uncertainty Assessment of Reservoir Simulation Models. Paper SPE-113390, Europec/EAGE Conference and Exhibition, Rome, Italy, 9-12 June.

Steiner, S., Ahsan, S. A., Noufal, A., Franco, B., Koksalan, T., Amjad, K., Helja, E., Alhosani, S., Adesanya, A., 2015. Integrated Approach to Evaluate Unconventional and Tight Reservoirs in Abu Dhabi. Paper SPE 177610-MS, presented at Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, UAE, 9-12 November.

Tavassoli, S., Xu, Y., and Sepehrnoori, K. 2018. Modeling Fault Reactivation Using Embedded Discrete Fracture Method. Paper SPE 191412, presented at SPE Annual Technical Conference and Exhibition, Dallas, Texas, 24-26 September.

Temizel, C., Balaji, K., Canbaz, C. H., Palabiyik, Y., Moreno, R., Rabiei, M., Ranjith, R. 2019. Data-Driven Analysis of Natural Gas EOR in Unconventional Shale Oils. Society of Petroleum Engineers. Paper SPE-195194, presented at SPE Oklahoma City Oil and Gas Symposium, Oklahoma City, Oklahoma, 9-10 April.

Vazquez, O., Young, C., Demyanov, V., Arnold, D., Fisher, A., MacMillan, A., and Christie, M. 2015. Produced-Water-Chemistry History Matching in the Janice Field. SPE Reservoir Evaluation & Engineering, 18(04), 564-576.

Wang, H., Echeverria-ciaurri, D., Durlofsky, L. J., and Cominelli, A. 2012. Optimal Well Placement Under Uncertainty Using A Retrospective Optimization Framework. SPE Journal 17(01): 112-121.

Wantawin, M., Yu, W., and Sepehrnoori, K. 2017a. An Iterative Work Flow for History Matching by Use of Design of Experiment, Response-Surface Methodology, and Markov Chain Monte Carlo Algorithm Applied to Tight Oil Reservoirs. SPE Reservoir Evaluation and Engineering 20 (3): 613-626.

Wantawin, M., Yu, W., Dachanuwattana, S., and Sepehrnoori, K. 2017b. An Iterative Response-Surface Methodology by Use of High-Degree-Polynomial Proxy Models for Integrated History Matching and Probabilistic Forecasting Applied to Shale-Gas Reservoirs. SPE J. 22 (6): 2012-2031.

Warren, J.E. and Root, P.J. 1963. The Behavior of Naturally Fractured Reservoirs. SPE Journal 3(3): 245-255.

Wu, P.-Y., Jain, V., Kulkarni, M. S., and Abubakar, A. 2018. Machine learning-based method for automated well-log processing and interpretation. Paper SEG-2018-2996973, presented at 2018 SEG International Exposition and Annual Meeting, Anaheim, California, 14-19 October.

Xu, Y., Cavalcante Filho, J.S.A., Yu, W., and Sepehrnoori, K. 2017a. Discrete-Fracture Modeling of Complex Hydraulic-Fracture Geometries in Reservoir Simulators. SPE Reservoir Evaluation and Engineering, 20(2): 403-422.

Xu, Y., Yu, W., and Sepehrnoori, K. 2017b. Modeling Dynamic Behaviors of Complex Fractures in Conventional Reservoir Simulators. Paper URTeC 2670513, presented at SPE/AAPG/SEG Unconventional Resources Technology Conference, Austin, Texas, 24-26 July.

Xu, Y., Yu, W., Li, N., Lolon, E., and Sepehrnoori, K. 2018. Modeling Well Performance in Piceance Basin Niobrara Formation using Embedded Discrete Fracture Model. Paper URTeC 2901327, presented at SPE/AAPG/SEG Unconventional Resources Technology Conference, Houston, Texas, 23-25 July.

Yin, J., Xie, J., Datta-Gupta A., and Hill, A.D. 2011. Improved Characterization and Performance Assessment of Shale Gas Wells by Integrating Stimulated Reservoir Volume and Production Data. Paper SPE 148969, presented at SPE Eastern Regional Meeting, Columbus, Ohio, 17 - 19 August.

Yu, W., and Sepehrnoori, K. 2014. Sensitivity Study and History Matching and Economic Optimization for Marcellus Shale. Paper URTEC-1923491, presented at

SPE/AAPG/SEG Unconventional Resources Technology Conference, Denver, Colorado, 25-27 August.

Yu, W., Tripoppoom, S., Sepehrnoori, K., and Miao, J. 2018c. An Automatic History-Matching Workflow for Unconventional Reservoirs Coupling MCMC and Non-Intrusive EDFM Methods. Society of Petroleum Engineers. Paper SPE-191473, presented at SPE Annual Technical Conference and Exhibition, Dallas, Texas, 24-26 September.

Yu, W., Xu, Y., Liu, M., Wu, K., and Sepehrnoori, K. 2018a. Simulation of Shale Gas Transport and Production with Complex Fractures using Embedded Discrete Fracture Model. AIChE Journal 64 (6): 2251-2264.

Yu, W., Xu, Y., Weijermars, R., Wu, K., and Sepehrnoori, K. 2018b. A Numerical Model for Simulating Pressure Response of Well Interference and Well Performance in Tight Oil Reservoirs with Complex-Fracture Geometries Using the Fast Embedded-Discrete-Fracture-Model Method. SPE Reservoir Evaluation and Engineering, 2(21): 489-502.

Yu, W., and Sepehrnoori, K. 2018. Shale Gas and Tight Oil Reservoir Simulation, 1st Ed.; Publisher: Elsevier, Cambridge, USA. ISBN: 978-0-12-813868-7.

Yu, W., Zhang, Y., Varavei, A., Sepehrnoori, K., Zhang, T., Wu, K., and Miao, J. 2019. Compositional Simulation of CO2 Huff-n-Puff in Eagle Ford Tight Oil Reservoirs with CO2 Molecular Diffusion, Nanopore Confinement and Complex Natural Fractures. SPE Reservoir Evaluation and Engineering, in preprint.

Zhang, Z., and Fassihi, M. R. 2013. Uncertainty Analysis and Assisted History Matching Workflow in Shale Condensate Reservoirs. Paper URTEC 1581398, presented at Unconventional Resources Technology Conference, Denver, 12-14 August.